Errors-and-Erasures Decoding for Block-Codes with Feedback

Barış Nakiboğlu
Electrical Engineering and Computer Science
Massachusetts Institute of Technology
MA 02139 Cambridge

Lizhong Zheng
Electrical Engineering and Computer Science
Massachusetts Institute of Technology
MA 02139 Cambridge

Abstract

Inner and outer bounds are derived on the optimal performance of fixed length block-codes on discrete memoryless channels with feedback and errors-and-erasures decoding. First an inner bound is derived using a two phase encoding scheme with communication and control phases together with the optimal decoding rule for the given encoding scheme, among decoding rules that can be represented in terms of pairwise comparisons between the messages. Then an outer bound is derived using a generalization of the straight-line bound to errors-and-erasures decoders and the optimal error exponent trade-off of a feedback encoder with two messages on a DMC. Finally upper and lower bounds are derived for the optimal erasure exponent of error free block-codes in terms of the rate.

I. Introduction:

Shannon showed in [28] that the capacity of discrete memoryless channels (DMCs) does not increase even when a noiseless and delay free feedback link is available from receiver to transmitter. On symmetric DMCs the sphere packing exponent bounds the error exponent of fixed length block-codes from above, as shown by Dobrushin[1] in [10]. Thus relaxations like errors-and-erasures decoding or variable length coding are needed for feedback to increase the error exponent of block-codes at rates larger than the critical rate on symmetric DMCs. In this work we investigate one such relaxation, namely errors-and-erasures decoding and find inner and outer bounds to the optimal error exponent erasure exponent tradeoff. This analysis complements the research on two related block coding schemes: variable length block coding and errors-and-erasures decoding for block-codes without feedback. We start with a very brief overview of the previous work on these problems to motivate our investigation.

Burnashev [3], [4], [5] was the first one to consider variable-length block-codes with feedback, instead of fixed length ones. He obtained the exact expression for the error exponent at all rates. Later Yamamoto and Itoh, [32], suggested a coding scheme which achieves the best error exponent for variable-length block-codes with feedback by using a fixed length block-code with an errors-and-erasures decoding, repetitively until a non-erasure decoding occurs.[2] In fact any fixed length block-code with erasures can be used repetitively, like it was done in [32], to get a variable length block-code with essentially the same error exponent as the original fixed length block-code. Thus [3] can be reinterpreted to give an upper bound to the error exponent achievable by fixed length block-codes with erasures. Furthermore this upper bound is achieved by the fixed length block-codes with erasures described [32], when erasure probability is decaying to zero sub-exponentially with block length. However the techniques used in this stream of work are insufficient for deriving proper inner or outer bounds for the situation when erasure probability is decaying exponentially with block length. As explained in the below paragraph the case with strictly positive erasure exponent is important both for engineering applications and for a better understanding of soft decoding with feedback. Our investigation provides proper tools for such an analysis, results in inner and outer bounds to the trade-off between error and erasure exponents, while recovering all previously known results for the zero erasure exponent case.

When considered together with higher layers, the codes in the physical layer are part of a variable length/delay communication scheme with feedback. However in the physical layer itself fixed length block-codes are used instead of variable length ones because of their amenability to modular design and robustness against the noise in the feedback link. In such an architecture retransmissions affects the performance of higher layers. The average transmission time

[1]Later Haroutunian, [16], established an upper bound on the error exponent of block-codes with feedback. This upper bound is equal to sphere packing exponent for symmetric channels but it is strictly larger than the sphere packing exponent for non-symmetric channels.

[2]Including erasures will not result in an increase in the exponent for variable-length block-codes with feedback.

is only a first order measure of this effect: as long as the erasure probability is vanishing with increasing block length, average transmission time will essentially be equal to the block length of the fixed length block-code. Thus with an analysis like the one in [32], the cost of retransmissions are ignored as long as the erasure probability goes to zero with increasing block length. In a communication system with multiple layers, however, retransmissions usually have costs beyond their effect on average transmission time, which are described by constraints on the probability distribution of the decoding time. Knowledge of error erasure exponent trade-off is useful in coming up with designs to meet those constraints. An example of this phenomena is variable length block coding schemes with a hard dead lines for decoding time, which has already been investigated by Gopala *et. al.* [15] for block-codes without feedback. They have used a block coding scheme with erasures and they have resend the message whenever an erasure occurs. But because of the hard dead line they employ this scheme only for some fixed number of trials. If all those trials fail, i.e. lead to an erasure, they use a non-erasure block-code. Using the error exponent erasure exponent trade-off they were able to obtain the best over all error performance for the given architecture.

This brings us to the second stream of research we complement with our investigation: errors-and-erasures decoding for block-codes without feedback. Forney [13] was the first one to consider errors-and-erasures decoding without feedback. He obtained an achievable trade-off between the exponents of error and erasure probabilities. Then Csiszár and Körner, [9] achieved the same performance using universal coding and decoding algorithms. Later Telatar and Gallager, [31], introduced a strict improvement on certain channels over the results presented in [13] and [9]. Recently there has been a revived interest in the errors and erasures decoding for universally achievable performances [21], [20], for alternative methods of analysis [19], for extensions to the channels with side information [25] and implementation with linear block-codes [17]. The encoding schemes in these codes do not have access to any feedback. However if the transmitter gets to learn whether the decoded message was an erasure or not, it can resend the message whenever it is erased. Because of this block retransmission variant these problems are sometimes called decision feedback problems.

We complement the results on the error exponent erasure exponent trade off without feedback and the results about error exponent of variable length block-codes with feedback, by finding inner and outer bounds to the error exponent erasure exponent trade off of fixed length block-codes with feedback. We first introduce our model and notation in the following Section II. Then in Section III we derive a lower bound using a two phase coding algorithm similar to the one described by Yamamoto and Ito in [32] and decoding rule and analysis techniques, inspired by Telatar's in [30] for the non-feedback case. Note that the analysis and the decoding rule in [30] is tailored for a single phase scheme and without feedback and the two phase scheme of [32] is tuned specifically to zero-erasure exponent; coming up with framework in which both of the ideas can be used efficiently is the main technical challenge here. In Section IV we first advance the straight line bound idea introduced by Shannon, Gallager and Berlekamp in [29] to block-codes with erasures. Then we use it together with an outer bound on the error exponent trade off between two codewords with feedback to establish an outer bounds. In Section V we first introduce error free block-codes with erasures and discuss its relation to the fixed length block-codes with errors and erasures, and then we present inner and outer bounds to the erasure exponent of error free block-codes and point out its relation to the error exponent erasure exponent trade off.

Before starting the presentation of our analysis, let us make a brief digression, and discuss two channel models in which the use of feedback had been investigated for block-codes without erasures. First channel model is the well known additive white Gaussian noise channel (AWGNC) model. In AWGNCs if the power constraint $\mathcal{P}$ is on the expected value of the energy spent on a block $\mathbf{E}[\mathcal{S}_{\mathsf{n}}]$ i.e. power constraint is of the form $\mathbf{E}[\mathcal{S}_{\mathsf{n}}] \leq \mathcal{P}\mathsf{n}$, the error probability can be made to decay faster than any exponential function with block length n. Schalkwijk and Kailath suggested a coding algorithm in [27] which achieves a doubly exponential decay in error probability for continuous time AWGNCs, i.e. infinite bandwidth case. Later Schalkwijk [26] modified that scheme to achieve the same performance in discrete time AWGNCs, i.e. finite bandwidth case. Concatenating Schalkwijk and Kailath scheme with pulse amplitude modulation stages, gives a multi-fold decrease in the error probability [24], [33], [14]. However this behavior relies on the non-existence of any amplitude limit, the particular form of the power constraint and the noise free nature of the feedback link. First of all, as observed in [5] and [22] when there is an amplitude limit, error probability decays exponentially with block length. More importantly if the power constraint restricts the energy spent in transmission of each message for all noise realizations, i.e. if the power constraint is an almost sure power

constraint[3] of the form $\mathcal{S}_{\mathsf{n}} \leq \mathcal{P}\mathsf{n}$; then sphere packing exponent is still an upper bound to the error exponent for AWGNCs as shown by Pinsker, [24]. Furthermore if the feedback link is also an AWGNC and if there is a power constraint[4] on the feedback transmissions, then even in the case when there are only two messages, error probability decays only exponentially as it has been recently shown by Kim *et.al.* [18].

The second channel model is the DMC model. Although feedback can not increase the error exponent for rates over the critical rate, it can simplify the encoding scheme [33], [12]. Furthermore, for rates below the critical rate it is possible to improve the error exponent using feedback. Zigangirov [33] has established lower bounds to the error exponent for BSCs using such a simple encoding scheme. Zigangirov's lower bound is equal to the sphere packing exponent for all rates in the interval $[R^{'}_{crit}, \mathcal{C}]$ where $R^{'}_{crit} < R_{crit}$ and Zigangirov's lower bound is strictly larger than the corresponding non-feedback exponent for rates below $R^{'}_{crit}$. Later Burnashev [6] has introduced an improvement to Zigangirov's bound for all positive rates less than $R^{'}_{crit}$. D'yachkov [12] generalized Zigangirov's encoding scheme for general DMC's and established a lower bounds to the error exponents for general binary input channels and k-ary symmetric channels. However it is still an open problem to find a constructive technique that can be used for all DMC's which outperforms the random coding bound. Like AWGNCs there has been a revived interest in the effect of a noisy feedback link and achievable performances with noisy feedback on DMCs. Burnashev and Yamamoto recently showed that error exponent of BSC channel increases even with a noisy feedback link [8], [7]. Furthermore Draper and Sahai [11] investigated the use of noisy feedback link in variable length schemes.

II. MODEL AND NOTATION:

The input and output alphabets of the forward channel are $\mathcal{X}$ and $\mathcal{Y}$, respectively. The channel input and output symbols at time t will be denoted by X_t and Y_t respectively. Furthermore, the sequences of input and output symbols from time t_1 to time t_2 are denoted by $\mathsf{X}_{t_1}^{t_2}$ and $\mathsf{Y}_{t_1}^{t_2}$. When $t_1 = 1$ we omit t_1 and simply write X^{t_2} and Y^{t_2} instead of $\mathsf{X}_1^{t_2}$ and $\mathsf{Y}_1^{t_2}$. The forward channel is a stationary memoryless channel characterized by an $|\mathcal{X}|$-by-$|\mathcal{Y}|$ transition probability matrix W.

$$\mathbf{P}\left[\mathsf{Y}_t | \, \mathsf{X}^t, \mathsf{Y}^{t-1}\right] = \mathbf{P}[\mathsf{Y}_t | \, \mathsf{X}_t] = W(\mathsf{Y}_t|\mathsf{X}_t) \qquad \forall t. \tag{1}$$

The feedback channel is noiseless and delay free, i.e. the transmitter observes Y_{t-1} before transmitting X_t.

The message M is drawn from the message set $\mathcal{M}$ with a uniform probability distribution and is given to the transmitter at time zero. At each time $t \in [1, \mathsf{n}]$ the input symbol $X_t(\mathsf{M}, \mathsf{Y}^{t-1})$ is sent. The sequence of functions $X_t(\cdot) : \mathcal{M} \times \mathcal{Y}^{t-1}$ which assigns an input symbol for each $m \in \mathcal{M}$ and $y^{t-1} \in \mathcal{Y}^{t-1}$ is called the encoding function.

After receiving Y^{n} the receiver decodes a $\hat{\mathsf{M}}(\mathsf{Y}^{\mathsf{n}}) \in \{\mathbf{x}\} \cup \mathcal{M}$ where $\mathbf{x}$ is the erasure symbol. The conditional erasure and error probabilities $P_{\mathbf{x}|\mathsf{M}}$ and $P_{\mathbf{e}|\mathsf{M}}$ and unconditional error and erasure probabilities, $P_{\mathbf{x}}$ and $P_{\mathbf{e}}$ are defined as,

$$P_{\mathbf{x}|\mathsf{M}} \triangleq \mathbf{P}\left[\hat{\mathsf{M}} = \mathbf{x} \middle| \mathsf{M}\right] \qquad P_{\mathbf{e}|\mathsf{M}} \triangleq \mathbf{P}\left[\hat{\mathsf{M}} \neq \mathsf{M} \middle| \mathsf{M}\right] - P_{\mathbf{x}|\mathsf{M}}$$

$$P_{\mathbf{x}} \triangleq \mathbf{P}\left[\hat{\mathsf{M}} = \mathbf{x}\right] \qquad P_{\mathbf{e}} \triangleq \mathbf{P}\left[\hat{\mathsf{M}} \neq \mathsf{M}\right] - P_{\mathbf{x}}$$

Since all the messages are equally likely we have,

$$P_{\mathbf{x}} = \tfrac{1}{|\mathcal{M}|} \sum\nolimits_m P_{\mathbf{x}|m} \qquad P_{\mathbf{e}} = \tfrac{1}{|\mathcal{M}|} \sum\nolimits_m P_{\mathbf{e}|m}$$

We use a somewhat abstract but rigorous approach in defining the rate and achievable exponent pairs. A reliable sequence $\mathcal{Q}$, is a sequence of codes indexed by their block lengths such that

$$\lim_{\mathsf{n} \to \infty} (P_{\mathbf{e}}{}^{(\mathsf{n})} + P_{\mathbf{x}}{}^{(\mathsf{n})} + \tfrac{1}{|\mathcal{M}^{(\mathsf{n})}|}) = 0.$$

In other words reliable sequences are sequences of codes whose overall error probability, detected and undetected, vanishes and whose size of message set diverges with block length n.

Definition 1: The rate, erasure exponent, and error exponent of a reliable sequence $\mathcal{Q}$ are given by

$$R_{\mathcal{Q}} \triangleq \liminf_{\mathsf{n} \to \infty} \tfrac{\ln |\mathcal{M}^{(\mathsf{n})}|}{\mathsf{n}} \qquad E_{\mathbf{x}\mathcal{Q}} \triangleq \liminf_{\mathsf{n} \to \infty} \tfrac{-\ln P_{\mathbf{x}}{}^{(\mathsf{n})}}{\mathsf{n}} \qquad E_{\mathbf{e}\mathcal{Q}} \triangleq \liminf_{\mathsf{n} \to \infty} \tfrac{-\ln P_{\mathbf{e}}{}^{(\mathsf{n})}}{\mathsf{n}}.$$

[3]As Kim *et. al.* [18] calls it.

[4]This constraint can be an expected or almost sure constraint.

Haroutunian, [16, Theorem 2], has already established a strong converse for erasure free block-codes with feedback which in our setting implies that $\lim_{\mathsf{n}\to\infty}({P_{\mathbf{e}}}^{(\mathsf{n})} + {P_{\mathbf{x}}}^{(\mathsf{n})}) = 1$ for all codes whose rates are strictly above the capacity, i.e. $R > \mathcal{C}$. Thus we consider only rates that are less than or equal to the capacity, $R \leq \mathcal{C}$. For all rates R below capacity and for all non-negative erasure exponents $E_{\mathbf{x}}$, we define the (true) error exponent $\mathcal{E}_{\mathbf{e}}(R, E_{\mathbf{x}})$ of fixed length block-codes with feedback to be the best error exponent of the reliable sequences[5] whose rate is at least R and whose erasure exponent is at least $E_{\mathbf{x}}$.

Definition 2: $\forall R \leq \mathcal{C}$ and $\forall E_{\mathbf{x}} \geq 0$ the error exponent, $\mathcal{E}_{\mathbf{e}}(R, E_{\mathbf{x}})$ is,

$$\mathcal{E}_{\mathbf{e}}(R, E_{\mathbf{x}}) \triangleq \sup_{\mathcal{Q}: R_{\mathcal{Q}} \geq R, E_{\mathbf{x}\mathcal{Q}} \geq E_{\mathbf{x}}} E_{\mathbf{e}\mathcal{Q}}. \tag{2}$$

Note that

$$\mathcal{E}_{\mathbf{e}}(R, E_{\mathbf{x}}) = \mathcal{E}(R) \qquad \forall E_{\mathbf{x}} > \mathcal{E}(R) \tag{3}$$

where $\mathcal{E}(R)$ is the (true) error exponent of erasure-free block-codes on DMCs with feedback.[6] Thus benefit of the errors-and-erasures decoding is the possible increase in the error exponent as the erasure exponent goes below $\mathcal{E}(R)$.

Determining $\mathcal{E}(R)$ for all R's and for all channels is still an open problem; only upper and lower bounds to $\mathcal{E}(R)$ are known. Our investigation focuses on quantifying the gains of errors-and-erasures decoding instead of finding $\mathcal{E}(R)$. Consequently, we restrict ourselves to the region where the erasure exponent is lower than the error exponent for the encoding scheme.

For future reference let us recall the expressions for the random coding exponent and the sphere packing exponent,

$$E_r(R, P) = \min_V \mathsf{D}\left(V \| W|P\right) + |\mathsf{I}\left(P, V\right) - R|^+ \qquad\qquad E_r(R) = \max_P E_r(R, P) \tag{4}$$

$$E_{sp}(R, P) = \min_{V: \mathsf{I}(P,V) \leq R} \mathsf{D}\left(V \| W|P\right) \qquad\qquad E_{sp}(R) = \max_P E_{sp}(R, P) \tag{5}$$

where $\mathsf{D}\left(V \| W|P\right)$ stands for conditional Kullback Leibler divergence of V and W under P, and $\mathsf{I}\left(P, V\right)$ stands for mutual information for input distribution P and channel V.

We denote the y marginal of a distribution like $P(x)V(y|x)$ by $(PV)_Y$. The support of a probability distribution P is denoted by $\mathsf{supp}P$

III. An Achievable Error Exponent - Erasure Exponent Trade Off

In this section we establish a lower bound to the achievable error exponent as a function of erasure exponent and rate. We use a two phase encoding scheme similar to the one described by Yamamoto and Ito in [32] together with a decoding rule similar to the one described by Telatar in [30]. In the first phase, the transmitter uses a fixed-composition code of length $\alpha\mathsf{n}$ and rate $\frac{R}{\alpha}$. At the end of the first phase, the receiver makes a maximum mutual information decoding to obtain a tentative decision $\tilde{\mathsf{M}}$. The transmitter knows $\tilde{\mathsf{M}}$ because of the feedback link. In $(\mathsf{n} - \mathsf{n}_1)$ long second phase the transmitter confirms the tentative decision by sending the accept codeword, if $\tilde{\mathsf{M}} = \mathsf{M}$, and rejects it by sending the reject codeword otherwise. At the end of the second phase the receiver either declares an erasure or declares the tentative decision as the decoded message. Receiver declares the tentative decision as the decoded message only when the tentative decision "dominates" all other messages. The word "dominate" will be made precise later in Section III-B. Our scheme is inspired by [32] and [30]. However, unlike [32] our decoding rule makes use of outputs of both of the phases instead of output of just second phase while deciding between declaring an erasure or declaring the tentative decision as the final one, and unlike [30] our encoding scheme is a feedback encoding scheme with two phases.

In the rest of this section, we analyze the performance of this coding architecture and derive the achievable error exponent expression in terms of a given rate R, erasure exponent $E_{\mathbf{x}}$, time sharing constant α, communication phase type P, control phase type (joint empirical type of the accept codeword and reject codeword) Π and domination rule $\succ$. Then we optimize over $\succ$, Π, P and α, to obtain an achievable error exponent expression as a function of rate R and erasure exponent $E_{\mathbf{x}}$.

[5]We restrict ourselves to the reliable sequences in order to ensure finite error exponent at zero erasure exponent. Note that a decoder which always declares erasures has zero erasure exponent and infinite error exponent.

[6]In order to see this consider a reliable sequence with erasures $\mathcal{Q}$ and replace its decoding algorithm by any erasure free one, $\mathcal{Q}'$ such that $\hat{\mathsf{M}}'(y^{\mathsf{n}}) = \hat{\mathsf{M}}(y^{\mathsf{n}})$ if $\hat{\mathsf{M}}(y^{\mathsf{n}}) \neq \mathbf{x}$. Then ${P_{\mathbf{e}}}^{(\mathsf{n})}_{\mathcal{Q}'} \leq {P_{\mathbf{x}}}^{(\mathsf{n})}_{\mathcal{Q}} + {P_{\mathbf{e}}}^{(\mathsf{n})}_{\mathcal{Q}}$; thus $E_{\mathbf{e}\mathcal{Q}'} = \min\{E_{\mathbf{x}\mathcal{Q}}, E_{\mathbf{e}\mathcal{Q}}\}$ and $R_{\mathcal{Q}'} = R_{\mathcal{Q}}$. This together with the definition of $\mathcal{E}(R)$ leads to equation (3).

A. Fixed-Composition Codes and The Packing Lemma

We start with a very brief overview of certain properties of types, a thorough handling of type idea can be found in [9]. The empirical distribution of an $x^{\mathsf{n}} \in \mathcal{X}^{\mathsf{n}}$ is called the type of x^{n} and the empirical distribution of transitions from a $x^{\mathsf{n}} \in \mathcal{X}^{\mathsf{n}}$ to a $y^{\mathsf{n}} \in \mathcal{X}^{\mathsf{n}}$ is called the conditional type:[7]

$$\mathsf{P}_{x^{\mathsf{n}}}(\tilde{x}) \triangleq \tfrac{1}{\mathsf{n}} \sum_{t=1}^{\mathsf{n}} \mathbb{1}_{\{x_t=\tilde{x}\}} \qquad \tilde{x} \in \mathcal{X}. \tag{6}$$

$$\mathsf{V}_{y^{\mathsf{n}}|x^{\mathsf{n}}}(\tilde{y}|\tilde{x}) \triangleq \tfrac{1}{\mathsf{n}\mathsf{P}_{x^{\mathsf{n}}}(\tilde{x})} \sum_{t=1}^{\mathsf{n}} \mathbb{1}_{\{x_t=\tilde{x}\}} \mathbb{1}_{\{y_t=\tilde{y}\}} \qquad \forall \tilde{y} \in \mathcal{Y}, \quad \forall \tilde{x} \text{ s.t. } \mathsf{P}_{x^{\mathsf{n}}}(\tilde{x}) > 0. \tag{7}$$

For any probability transition matrix $W : \mathsf{supp}\mathsf{P}_{x^{\mathsf{n}}} \to \mathcal{Y}$ we have[8]

$$\prod_{t=1}^{\mathsf{n}} W(y_t|x_t) = e^{-\mathsf{n}(\mathsf{D}(\mathsf{V}_{y^{\mathsf{n}}|x^{\mathsf{n}}} \| W|\mathsf{P}_{x^{\mathsf{n}}}) + \mathsf{H}(\mathsf{V}_{y^{\mathsf{n}}|x^{\mathsf{n}}}|\mathsf{P}_{x^{\mathsf{n}}}))} \tag{8}$$

The set of all y^{n}'s with the same conditional type V with respect to x^{n} is called the V-shell of x^{n} and denoted by $T_V(x^{\mathsf{n}})$:

$$T_V(x^{\mathsf{n}}) = \{y^{\mathsf{n}} : \mathsf{V}_{y^{\mathsf{n}}|x^{\mathsf{n}}} = V\}. \tag{9}$$

Note that for any transition probabilities from $\mathcal{X}$ to $\mathcal{Y}$ total probability of $T_V(x^{\mathsf{n}})$ has to be less than one. Thus by assuming that transition probabilities are V and using equation (8) we can conclude that,

$$|T_V(x^{\mathsf{n}})| \le e^{\mathsf{H}(\mathsf{V}_{y^{\mathsf{n}}|x^{\mathsf{n}}}|\mathsf{P}_{x^{\mathsf{n}}}))} \tag{10}$$

Codes whose codewords all have the same empirical distribution, $\mathsf{P}_{x^{\mathsf{n}}(m)} = P \ \forall m \in \mathcal{M}$ are called fixed-composition codes. In Section III-D we will describe the error and erasure events in terms of the intersections of $V-$shells of different codewords. For doing that let us define $F^{(\mathsf{n})}\left(V, \hat{V}, m\right)$ as the intersection of V-shell of $x^{\mathsf{n}}(m)$ and the $\hat{V}$-shells of other codewords:

$$F^{(\mathsf{n})}\left(V, \hat{V}, m\right) \triangleq T_V(x^{\mathsf{n}}(m)) \bigcap \cup_{\tilde{m} \ne m} T_{\hat{V}}(x^{\mathsf{n}}(\tilde{m})). \tag{11}$$

The following packing lemma, proved by Csiszár and Körner [9, Lemma 2.5.1], claims the existence of a code with a guaranteed upper bound on the size of $F^{(\mathsf{n})}\left(V, \hat{V}, m\right)$.

Lemma 1: For every block length $\mathsf{n} \ge 1$, rate $R > 0$ and type P satisfying $H(P) > R$, there exist at least $\lfloor e^{\mathsf{n}(R-\delta_{\mathsf{n}})} \rfloor$ distinct type P sequences in $\mathcal{X}^{\mathsf{n}}$ such that for every pair of stochastic matrices $V : \mathsf{supp}P \to \mathcal{Y}$, $\hat{V} : \mathsf{supp}P \to \mathcal{Y}$ and $\forall m \in \mathcal{M}$

$$\left| F^{(\mathsf{n})}\left(V, \hat{V}, m\right) \right| \le |T_V(x^{\mathsf{n}}(m))| e^{-\mathsf{n}|I(P,\hat{V})-R|^{+}}$$

where $\delta_{\mathsf{n}} = \frac{\ln 4 + (4|\mathcal{X}| + 6|\mathcal{X}||\mathcal{Y}|)\ln(\mathsf{n}+1)}{\mathsf{n}}$.

Above lemma is stated in a slightly different way by the authors of [9], for a fixed δ and large enough n. However, this form follows immediately from their proof.

If we use Lemma 1 together with equations (8) and (10) we can bound the conditional probability of observing a $y^{\mathsf{n}} \in F^{(\mathsf{n})}\left(V, \hat{V}, m\right)$ when $\mathsf{M} = m$ as follows.

Corollary 1: In a code satisfying Lemma 1, when message $m \in \mathcal{M}$ is sent, the probability of getting a $y^{\mathsf{n}} \in T_V(x^{\mathsf{n}}(m))$ which is also in $T_{\hat{V}}(x^{\mathsf{n}}(\tilde{m}))$, for some $\tilde{m} \in \mathcal{M}$ such that $\tilde{m} \ne m$ is bounded as follows,

$$\mathbf{P}\left[F^{(\mathsf{n})}\left(V, \hat{V}, \mathsf{M}\right) \middle| \mathsf{M} \right] \le e^{-\mathsf{n}\eta\left(R,P,V,\hat{V}\right)} \tag{12}$$

where

$$\eta\left(R, P, V, \hat{V}\right) \triangleq \mathsf{D}(V \| W|P) + \left|\mathsf{I}\left(P, \hat{V}\right) - R\right|^{+} \tag{13}$$

[7]Note that $\mathsf{P}_{y^{\mathsf{n}}}$ corresponds to a distribution on $\mathcal{X}$ for all $x^{\mathsf{n}} \in \mathcal{X}^{\mathsf{n}}$, where as $\mathsf{V}_{y^{\mathsf{n}}|x^{\mathsf{n}}}$ determines a channel from the support of $\mathsf{P}_{x^{\mathsf{n}}}$ to $\mathcal{Y}$.

[8]Note that for any $W : \mathcal{X} \to \mathcal{Y}$ there is unique consistent $W' : \mathsf{supp}\mathsf{P}_{x^{\mathsf{n}}} \to \mathcal{Y}$.

B. Coding Algorithm

In the first phase, the communication phase, we use a length $\mathsf{n}_1 = \lceil \alpha \mathsf{n} \rceil$ type P fixed-composition code with $\lfloor e^{\mathsf{n}_1(\frac{R}{\alpha} - \delta_{\mathsf{n}_1})} \rfloor$ codewords which satisfies the property described in Lemma 1. At the end of the first phase the receiver makes a tentative decision by choosing the codeword that has the maximum empirical mutual information with the output sequence $\mathsf{Y}^{\mathsf{n}_1}$. If there is a tie, i.e. if there are more than one codewords which have maximum empirical mutual information, the receiver chooses the codeword which has the lowest index.

$$\tilde{\mathsf{M}} = \left\{ m : \begin{array}{ll} \mathsf{I}\left(P, \mathsf{V}_{\mathsf{Y}^{\mathsf{n}_1}|x^{\mathsf{n}}(m)}\right) > \mathsf{I}\left(P, \mathsf{V}_{\mathsf{Y}^{\mathsf{n}_1}|x^{\mathsf{n}}(\tilde{m})}\right) & \forall \tilde{m} < m \\ \mathsf{I}\left(P, \mathsf{V}_{\mathsf{Y}^{\mathsf{n}_1}|x^{\mathsf{n}}(m)}\right) \geq \mathsf{I}\left(P, \mathsf{V}_{\mathsf{Y}^{\mathsf{n}_1}|x^{\mathsf{n}}(\tilde{m})}\right) & \forall \tilde{m} > m \end{array} \right\} \tag{14}$$

In the remaining $(\mathsf{n} - \mathsf{n}_1)$ time units, the transmitter sends the accept codeword $x^{\mathsf{n}}_{\mathsf{n}_1+1}(a)$ if $\tilde{\mathsf{M}} = \mathsf{M}$ and sends the reject codeword $x^{\mathsf{n}}_{\mathsf{n}_1+1}(r)$ otherwise.

Note that our encoding scheme uses the feedback link actively for the encoding neither within the first phase nor within the second phase. It does not even change the codewords it uses for accepting or rejecting the tentative decision depending on the observation in the first phase. Feedback is only used to reveal the tentative decision to the transmitter.

Accept and reject codewords have joint type $\Pi(\tilde{x}, \tilde{\tilde{x}})$, i.e. the ratio of the number of time instances in which accept codeword has an $\tilde{x} \in \mathcal{X}$ and reject codeword has a $\tilde{\tilde{x}} \in \mathcal{X}$ to the length of the codewords, $(\mathsf{n} - \mathsf{n}_1)$, is $\Pi(\tilde{x}, \tilde{\tilde{x}})$. The joint conditional type of the output sequence in the second phase, $\mathsf{U}_{y^{\mathsf{n}}_{\mathsf{n}_1+1}}$, is the empirical conditional distribution of $y^{\mathsf{n}}_{\mathsf{n}_1+1}$. We call set of all output sequences $y^{\mathsf{n}}_{\mathsf{n}_1+1}$ whose joint conditional type is U, the U-shell and denote it by T_U.

Like we did in the Corollary 1, we can upper bound the probability of U-shells. Note that if $\mathsf{Y}^{\mathsf{n}}_{\mathsf{n}_1+1} \in T_U$ then,

$$\mathbf{P}\left[\mathsf{Y}^{\mathsf{n}}_{\mathsf{n}_1+1}\middle|\, \mathsf{X}^{\mathsf{n}}_{\mathsf{n}_1+1} = x^{\mathsf{n}}_{\mathsf{n}_1+1}(a)\right] = e^{-(\mathsf{n}-\mathsf{n}_1)(\mathsf{D}(U\|W_a|\Pi) + \mathsf{H}(U|\Pi))}$$
$$\mathbf{P}\left[\mathsf{Y}^{\mathsf{n}}_{\mathsf{n}_1+1}\middle|\, \mathsf{X}^{\mathsf{n}}_{\mathsf{n}_1+1} = x^{\mathsf{n}}_{\mathsf{n}_1+1}(r)\right] = e^{-(\mathsf{n}-\mathsf{n}_1)(\mathsf{D}(U\|W_r|\Pi) + \mathsf{H}(U|\Pi))}$$

where $x^{\mathsf{n}}_{\mathsf{n}_1+1}(a)$ is the accept codeword, $x^{\mathsf{n}}_{\mathsf{n}_1+1}(r)$ is the reject codeword, $W_a(y|\tilde{x}, \tilde{\tilde{x}}) = W(y|\tilde{x})$ and $W_r(y|\tilde{x}, \tilde{\tilde{x}}) = W(y|\tilde{\tilde{x}})$. Noting that $|T_U| \leq e^{-(\mathsf{n}-\mathsf{n}_1)\mathsf{H}(U|\Pi)}$, we get:

$$\mathbf{P}\left[T_U|\, \mathsf{X}^{\mathsf{n}}_{\mathsf{n}_1+1} = x^{\mathsf{n}}_{\mathsf{n}_1+1}(a)\right] \leq e^{-(\mathsf{n}-\mathsf{n}_1)\mathsf{D}(U\|W_a|\Pi)} \tag{15a}$$
$$\mathbf{P}\left[T_U|\, \mathsf{X}^{\mathsf{n}}_{\mathsf{n}_1+1} = x^{\mathsf{n}}_{\mathsf{n}_1+1}(r)\right] \leq e^{-(\mathsf{n}-\mathsf{n}_1)\mathsf{D}(U\|W_r|\Pi)}. \tag{15b}$$

C. Decoding Rule

For an encoder like the one in Section III-B, a decoder that depends only on the conditional type of $\mathsf{Y}^{\mathsf{n}_1}$ for different codewords in the communication phase, i.e. $\mathsf{V}_{\mathsf{Y}^{\mathsf{n}_1}|x^{\mathsf{n}_1}(m)}$'s for $m \in \mathcal{M}$, the conditional type of the channel output in the control phase, i.e. $\mathsf{U}_{\mathsf{Y}^{\mathsf{n}}_{\mathsf{n}_1+1}}$, and the indices of the codewords can achieve the minimum error probability for a given erasure probability. However finding that decoder becomes analytically intractable problem early on. Instead, we restrict ourselves to the decoders that can be written in terms of pair wise comparisons between messages given Y^{n}. Furthermore we assumes that these pairwise comparisons depend only on the conditional type of $\mathsf{Y}^{\mathsf{n}_1}$ for the messages compared, the conditional output type in the control phase and the indices of the messages. Thus if the triplet corresponding to the tentative decision $(\mathsf{V}_{\mathsf{Y}^{\mathsf{n}_1}|x^{\mathsf{n}_1}(\tilde{\mathsf{M}})}, \mathsf{U}_{\mathsf{Y}^{\mathsf{n}}_{\mathsf{n}_1+1}}, \tilde{\mathsf{M}})$ dominates all other triplets of the form $(\mathsf{V}_{\mathsf{Y}^{\mathsf{n}_1}|x^{\mathsf{n}_1}(m)}, \mathsf{U}_{\mathsf{Y}^{\mathsf{n}}_{\mathsf{n}_1+1}}, m)$ for $m \neq \tilde{\mathsf{M}}$, the tentative decision becomes final; else an erasure is declared.[9]

$$\hat{\mathsf{M}} = \begin{cases} \tilde{\mathsf{M}} & \text{if } \forall m \neq \tilde{\mathsf{M}} \quad (\mathsf{V}_{\mathsf{Y}^{\mathsf{n}_1}|\tilde{\mathsf{M}}}, \mathsf{U}_{\mathsf{Y}^{\mathsf{n}}_{\mathsf{n}_1+1}}, \tilde{\mathsf{M}}) \succ (\mathsf{V}_{\mathsf{Y}^{\mathsf{n}_1}|m}, \mathsf{U}_{\mathsf{Y}^{\mathsf{n}}_{\mathsf{n}_1+1}}, m) \\ \mathbf{x} & \text{if } \exists m \neq \tilde{\mathsf{M}} \text{ s.t. } (\mathsf{V}_{\mathsf{Y}^{\mathsf{n}_1}|\tilde{\mathsf{M}}}, \mathsf{U}_{\mathsf{Y}^{\mathsf{n}}_{\mathsf{n}_1+1}}, \tilde{\mathsf{M}}) \not\succ (\mathsf{V}_{\mathsf{Y}^{\mathsf{n}_1}|m}, \mathsf{U}_{\mathsf{Y}^{\mathsf{n}}_{\mathsf{n}_1+1}}, m) \end{cases} \tag{16}$$

The binary relation $\succ$ is such that if (V, U, m) dominates $(\hat{V}, U, \tilde{m})$ then $(\hat{V}, U, \tilde{m})$ does not dominate (V, U, m):

$$(V, U, m) \succ (\hat{V}, U, \tilde{m}) \Rightarrow (\hat{V}, U, \tilde{m}) \not\succ (V, U, m).$$

This property is a necessary and sufficient condition for a binary relation to be a domination rule. Decoder given (16), however, either accepts or rejects the tentative decision $\tilde{\mathsf{M}}$ given in (14). Consequently its domination rule also satisfies following two properties:

[9]Note that conditional probability, $\mathbf{P}[\mathsf{Y}^{\mathsf{n}}|\, \mathsf{M} = m]$, is only a function of corresponding $\mathsf{V}_{\mathsf{Y}^{\mathsf{n}_1}|x^{\mathsf{n}}(m)}$ and $\mathsf{U}_{\mathsf{Y}^{\mathsf{n}}_{\mathsf{n}_1+1}}$. Thus all decoding rules, that accepts or rejects the tentative decision, $\tilde{\mathsf{M}}$, based on a threshold test on likelihood ratios, $\frac{\mathbf{P}[\mathsf{Y}^{\mathsf{n}}|\mathsf{M}=\tilde{\mathsf{M}}]}{\mathbf{P}[\mathsf{Y}^{\mathsf{n}}|\mathsf{M}=m]}$, for $m \neq \tilde{\mathsf{M}}$ are in this family of decoding rules.

(a) If the empirical mutual information of the messages in the communication phase are not equal, only the message with larger mutual information can dominate the other one.
(b) If the empirical mutual information of the messages in the communication phase are equal, only the message with lower index can dominate the other one.

For any such binary relation there is a corresponding decoder of the form given in Equation (16). In our scheme we either use the trivial domination rule leading to the trivial decoder $\hat{\mathsf{M}} = \tilde{\mathsf{M}}$ or the domination rule given in equation (17), both of which satisfies these conditions.

$$(V, U, m) \succ (\hat{V}, U, \tilde{m}) \Leftrightarrow \begin{cases} \mathsf{I}(P, V) > \mathsf{I}\left(P, \hat{V}\right) \text{ and } \alpha\eta\left(\frac{R}{\alpha}, P, V, \hat{V}\right) + (1-\alpha)\mathsf{D}\left(U \| W_a|\Pi\right) \leq E_{\mathbf{x}} & \text{if } m \geq \tilde{m} \\ \mathsf{I}(P, V) \geq \mathsf{I}\left(P, \hat{V}\right) \text{ and } \alpha\eta\left(\frac{R}{\alpha}, P, V, \hat{V}\right) + (1-\alpha)\mathsf{D}\left(U \| W_a|\Pi\right) \leq E_{\mathbf{x}} & \text{if } m < \tilde{m} \end{cases} \tag{17}$$

where $\eta\left(R, P, V, \hat{V}\right)$ is given by the equation (13).

Among the family of decoders we are considering, i.e. among the decoders that only depend on the pairwise comparisons between conditional types and indices of the messages compared, the decoder given in (16) and (17) is optimal in terms of error-exponent-erasure-exponent tradeoff. Furthermore, in order to employ this decoding rule, the receiver needs to determine only the two messages with the highest empirical mutual information in the first phase. Then the receiver needs to check whether the triplet corresponding to the tentative decision dominates the triplet corresponding to the message with the second highest empirical mutual information. If it does then, for the rule given in (17), it is guaranteed to dominate rest of the triplets too.

D. Error Analysis

Using an encoder like the one described in Section III-B and a decoder like the one in (16) we achieve the performance given below. If $E_{\mathbf{x}} \leq \alpha E_r(\frac{R}{\alpha}, P)$ then the domination rule given in equation (17) is used in the decoder; else a trivial domination rule that leads to a non-erasure decoder, $\hat{\mathsf{M}} = \tilde{\mathsf{M}}$, is used in the decoder.

Theorem 1: For any block length $\mathsf{n} \geq 1$, rate R, erasure exponent $E_{\mathbf{x}}$, time sharing constant α, communication phase type P and control phase type Π, there exists a length n block-code with feedback such that

$$\ln|\mathcal{M}| \geq e^{\mathsf{n}(R-\delta_{\mathsf{n}})} \qquad P_{\mathbf{x}} \leq e^{-\mathsf{n}(E_{\mathbf{x}}-\delta'_{\mathsf{n}})} \qquad P_{\mathbf{e}} \leq e^{-\mathsf{n}(E_{\mathbf{e}}(R,E_{\mathbf{x}},\alpha,P,\Pi)-\delta'_{\mathsf{n}})}$$

where $E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$ is given by,

$$E_{\mathbf{e}} = \begin{cases} \alpha E_r(\frac{R}{\alpha}, P) & \text{if } E_{\mathbf{x}} > \alpha E_r(\frac{R}{\alpha}, P) \\ \min\limits_{\substack{(V,\hat{V},U):(V,\hat{V},U)\in\mathcal{V} \\ \alpha\eta\left(\frac{R}{\alpha},P,V,\hat{V}\right)+(1-\alpha)\mathsf{D}\left(U\|W_a|\Pi\right)\leq E_{\mathbf{x}}}} \alpha\eta\left(\frac{R}{\alpha}, P, \hat{V}, V\right) + (1-\alpha)\mathsf{D}\left(U\| W_r|\Pi\right) & \text{if } E_{\mathbf{x}} \leq \alpha E_r(\frac{R}{\alpha}, P) \end{cases} \tag{18a}$$

$$\mathcal{V} = \{(V_1, V_2, U) : \mathsf{I}(P, V_1) \geq \mathsf{I}(P, V_2) \text{ and } (PV_1)_Y = (PV_2)_Y\} \tag{18b}$$

$$\delta'_{\mathsf{n}} = \tfrac{(|\mathcal{X}|+1)^2|\mathcal{Y}|\log(\mathsf{n}+1)}{\mathsf{n}} \tag{18c}$$

The optimization problem given in (18) is a convex optimization problem: it is minimization of a convex function over a convex set. Thus the value of the exponent, $E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$ can numerically be calculated relatively easily. Furthermore $E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$ can be written in terms of solutions of lower dimensional optimization problems (see equation (37). However problem of finding the optimal (α, P, Π) triple for a given $(R, E_{\mathbf{x}})$ pair is not that easy in general, as we will discuss in more detail in Section III-E.

Note that for all control phase types Π and control phase output types U, $\mathsf{D}\left(U\| W_a|\Pi\right) \geq 0$, $\mathsf{D}\left(U\| W_r|\Pi\right) \geq 0$. Using this fact together with the definitions of $E_r(R, P)$ and $\eta\left(R, P, \hat{V}, V\right)$ given in (4) and (13) we get:

$$E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi) \geq \alpha E_r(\tfrac{R}{\alpha}, P) \qquad \forall (R, E_{\mathbf{x}}, \alpha, P, \Pi) \text{ s.t. } E_{\mathbf{x}} \leq \alpha E_r(\tfrac{R}{\alpha}, P) \tag{19}$$

Since we are interested in quantifying the gains of errors-and-erasures decoding over the decoding schemes without erasures we are ultimately interested only in the region where $E_{\mathbf{x}} \leq \alpha E_r(\frac{R}{\alpha}, P)$ holds. However equation (18) gives us the whole achievable region for the family of codes we are considering.

Proof: A decoder of the form given in (16) decodes correctly when $\tilde{\mathsf{M}} = \mathsf{M}$ and $(\mathsf{Y}^{\mathsf{n}}, \mathsf{M}) \succ (\mathsf{Y}^{\mathsf{n}}, m)$ for all[10] $m \neq \mathsf{M}$. Thus an error or an erasure occur only when the correct message does not dominate all other messages, i.e. when $\exists m \neq \mathsf{M}$ such that $(\mathsf{Y}^{\mathsf{n}}, \mathsf{M}) \not\succ (\mathsf{Y}^{\mathsf{n}}, m)$. Consequently, we can write the sum of conditional error and erasure probabilities for a message $m \in \mathcal{M}$ as,

$$P_{\mathbf{e}|m} + P_{\mathbf{x}|m} = \mathbf{P}[\{y^{\mathsf{n}} : \exists \tilde{m} \neq m \text{ s.t.}(y^{\mathsf{n}}, m) \not\succ (y^{\mathsf{n}}, \tilde{m})\} | \, \mathsf{M} = m] \tag{20}$$

This can happen in two ways, either there is an error in the first phase, i.e. $\tilde{\mathsf{M}} \neq m$ or first phase tentative decision is correct, i.e. $\tilde{\mathsf{M}} = m$, but the second phase observation $y_{\mathsf{n}_1+1}^{\mathsf{n}}$ leads to an erasure i.e. $\hat{\mathsf{M}} = \mathbf{x}$. For a decoder using a domination rule satisfying constraints described in Section III-C,

$$\begin{aligned} P_{\mathbf{e}|m} + P_{\mathbf{x}|m} \leq & \sum_{V} \sum_{\hat{V}: \mathsf{I}(P,\hat{V}) \geq \mathsf{I}(P,V)} \sum_{y^{\mathsf{n}_1} \in F^{(\mathsf{n}_1)}(V,\hat{V},m)} \mathbf{P}[y^{\mathsf{n}_1} | \, m] \\ & + \sum_{V} \sum_{\hat{V}: \mathsf{I}(P,\hat{V}) \leq \mathsf{I}(P,V)} \sum_{y^{\mathsf{n}_1} \in F^{(\mathsf{n}_1)}(V,\hat{V},m)} \mathbf{P}[y^{\mathsf{n}_1} | \, m] \sum_{U:(V,U,m) \not\succ (\hat{V},U,m+1)} \sum_{y_{\mathsf{n}_1+1}^{\mathsf{n}} \in T_U} \mathbf{P}\left[y_{\mathsf{n}_1+1}^{\mathsf{n}} \middle| \, x_{\mathsf{n}_1+1}^{\mathsf{n}}(a)\right]. \end{aligned}$$

where[11] $F^{(\mathsf{n}_1)}\left(V, \hat{V}, m\right)$ is the intersection of V-shell of message $m \in \mathcal{M}$ with the $\hat{V}$-shells of other messages, defined in equation (11). As result of Corollary 1,

$$\begin{aligned} \sum_{y^{\mathsf{n}_1} \in F^{(\mathsf{n}_1)}(V,\hat{V},m)} \mathbf{P}[y^{\mathsf{n}_1} | \, m] &= \mathbf{P}\left[F^{(\mathsf{n}_1)}\left(V, \hat{V}, m\right) \middle| \, \mathsf{M} = m\right] \\ &\leq e^{-\mathsf{n}_1 \eta\left(\frac{R}{\alpha}, P, V, \hat{V}\right)}. \end{aligned}$$

Furthermore, as result of equation (15a) we have

$$\begin{aligned} \sum_{y_{\mathsf{n}_1+1}^{\mathsf{n}} \in T_U} \mathbf{P}\left[y_{\mathsf{n}_1+1}^{\mathsf{n}} \middle| \, x_{\mathsf{n}_1+1}^{\mathsf{n}}(a)\right] &= \mathbf{P}\left[T_U | \, \mathsf{X}_{\mathsf{n}_1+1}^{\mathsf{n}} = x_{\mathsf{n}_1+1}^{\mathsf{n}}(a)\right] \\ &\leq e^{-(\mathsf{n}-\mathsf{n}_1)\mathsf{D}(U \| W_a | \Pi)}. \end{aligned}$$

In addition the number of different non-empty V-shells in the communication phase is less than $(\mathsf{n}_1+1)^{|\mathcal{X}||\mathcal{Y}|}$ and the number of non-empty U-shells in the control phase is less than $(\mathsf{n}-\mathsf{n}_1+1)^{|\mathcal{X}|^2|\mathcal{Y}|}$. We denote the set of $(V, \hat{V}, U)$ triples that corresponds to erasures with a correct tentative decision by $\mathcal{V}_{\mathbf{x}}$:

$$\mathcal{V}_{\mathbf{x}} \triangleq \left\{(V, \hat{V}, U) : \mathsf{I}\,(P, V) \geq \mathsf{I}\left(P, \hat{V}\right) \text{ and } (PV)_Y = (P\hat{V})_Y \text{ and } (V, U, m) \not\succ (\hat{V}, U, m+1)\right\}. \tag{21}$$

In the above definition m is a dummy variable and $\mathcal{V}_{\mathbf{x}}$ is the same set for all $m \in \mathcal{M}$. Thus using (21) we get

$$\begin{aligned} P_{\mathbf{e}|m} + P_{\mathbf{x}|m} \leq & (\mathsf{n}_1+1)^{2|\mathcal{X}||\mathcal{Y}|} \max_{V,\hat{V}: \mathsf{I}(P,V) \leq \mathsf{I}(P,\hat{V})} e^{-\mathsf{n}_1 \eta\left(R/\alpha, P, V, \hat{V}\right)} \\ & + (\mathsf{n}_1+1)^{2|\mathcal{X}||\mathcal{Y}|} (\mathsf{n}-\mathsf{n}_1+1)^{|\mathcal{X}|^2|\mathcal{Y}|} \max_{(V,\hat{V},U) \in \mathcal{V}_{\mathbf{x}}} e^{-\mathsf{n}_1(\eta\left(R/\alpha, P, V, \hat{V}\right)) + (\mathsf{n}-\mathsf{n}_1)\mathsf{D}(U \| W_a | \Pi))}. \end{aligned}$$

Using the definition of $E_r(\frac{R}{\alpha}, P)$ given in (4) we get

$$P_{\mathbf{e}|m} + P_{\mathbf{x}|m} \leq e^{\mathsf{n}\delta'_{\mathsf{n}}} \max\left\{e^{-\mathsf{n}\alpha E_r(R/\alpha, P)}, e^{-\mathsf{n} \min_{(V,\hat{V},U) \in \mathcal{V}_{\mathbf{x}}} \alpha\eta\left(R/\alpha, P, V, \hat{V}\right) + (1-\alpha)\mathsf{D}(U \| W_a | \Pi)}\right\}. \tag{22}$$

On the other hand an error occurs only when an incorrect message dominates all other messages, i.e. when $\exists \tilde{m} \neq m$ such that $(\mathsf{Y}^{\mathsf{n}}, \tilde{m}) \succ (\mathsf{Y}^{\mathsf{n}}, \tilde{\tilde{m}})$ for all $\tilde{m} \neq \tilde{\tilde{m}}$:

$$P_{\mathbf{e}|m} = \mathbf{P}\left[\left\{y^{\mathsf{n}} : \exists \tilde{m} \neq m \text{ s.t. } (y^{\mathsf{n}}, \tilde{m}) \succ (y^{\mathsf{n}}, \tilde{\tilde{m}}) \quad \forall \tilde{\tilde{m}} \neq \tilde{m}\right\} \middle| \, \mathsf{M} = m\right].$$

Note that when a $\tilde{m} \in \mathcal{M}$ dominates all other $\tilde{\tilde{m}} \neq \tilde{m}$, it also dominates m, i.e.

$$\left\{y^{\mathsf{n}} : \exists \tilde{m} \neq m \text{ s.t.}(y^{\mathsf{n}}, \tilde{m}) \succ (y^{\mathsf{n}}, \tilde{\tilde{m}}) \quad \forall \tilde{\tilde{m}} \neq \tilde{m}\right\} \subset \{y^{\mathsf{n}} : \exists \tilde{m} \neq m \text{ s.t.}(y^{\mathsf{n}}, \tilde{m}) \succ (y^{\mathsf{n}}, m)\}.$$

[10]We use the short hand $(\mathsf{Y}^{\mathsf{n}}, \mathsf{M}) \succ (\mathsf{Y}^{\mathsf{n}}, m)$ for $(\mathsf{V}_{\mathsf{Y}^{\mathsf{n}_1}|\mathsf{M}}, \mathsf{U}_{\mathsf{Y}_{\mathsf{n}_1+1}^{\mathsf{n}}}, \mathsf{M}) \succ (\mathsf{V}_{\mathsf{Y}^{\mathsf{n}_1}|m}, \mathsf{U}_{\mathsf{Y}_{\mathsf{n}_1+1}^{\mathsf{n}}}, m)$ in the rest of this section.

[11]Note that for the case when $m = |\mathcal{M}|$, we need to replace $(V, U, m) \not\succ (\hat{V}, U, m+1)$ with $(V, U, m-1) \not\succ (\hat{V}, U, m)$.

Thus,

$$
\begin{aligned}
P_{\mathbf{e}|m} &\leq \mathbf{P}[\{y^{\mathsf{n}} : \exists \tilde{m} \neq m \text{ s.t.} (y^{\mathsf{n}}, \tilde{m}) \succ (y^{\mathsf{n}}, m)\} | \, \mathsf{M} = m] \\
&= \sum_{V} \sum_{\hat{V}: \mathsf{I}\left(P, \hat{V}\right) \geq \mathsf{I}(P,V)} \sum_{y^{\mathsf{n}_1} \in F^{(\mathsf{n}_1)}\left(V, \hat{V}, m\right)} \mathbf{P}[y^{\mathsf{n}_1} | \, \mathsf{M} = m] \sum_{U: (\hat{V}, U, m-1) \succ (V, U, m)} \sum_{y^{\mathsf{n}}_{\mathsf{n}_1+1} \in T_U} \mathbf{P}\left[y^{\mathsf{n}}_{\mathsf{n}_1+1} \middle| \, x^{\mathsf{n}}_{\mathsf{n}_1+1}(r)\right]. \qquad (23)
\end{aligned}
$$

The tentative decision is not equal to m only if there is a message with a strictly higher empirical mutual information or if there is a messages which has equal mutual information but smaller index. This is the reason why we sum over $(\hat{V}, U, m-1) \succ (V, U, m)$. Using the inequality (15b) in the inner most two sums and then applying inequality (12) we get,

$$
\begin{aligned}
P_{\mathbf{e}|m} &\leq (\mathsf{n}+1)^{(|\mathcal{X}|^2 + 2|\mathcal{X}|)|\mathcal{Y}|} \max_{\substack{(V, \hat{V}, U): \, \mathsf{I}\left(P, \hat{V}\right) \geq \mathsf{I}(P,V) \\ (\hat{V}, U, m-1) \succ (V, U, m)}} e^{-\mathsf{n}\left(\alpha\eta\left(R/\alpha, P, V, \hat{V}\right) + (1-\alpha)\mathsf{D}(U \| W_r | \Pi)\right)} \\
&\leq e^{\mathsf{n}\delta'_{\mathsf{n}}} e^{-\mathsf{n} \min_{(\hat{V}, V, U) \in \mathcal{V}_{\mathbf{e}}} \left(\alpha\eta\left(R/\alpha, P, V, \hat{V}\right) + (1-\alpha)\mathsf{D}(U \| W_r | \Pi)\right)} \\
&= e^{\mathsf{n}\delta'_{\mathsf{n}}} e^{-\mathsf{n} \min_{(V, \hat{V}, U) \in \mathcal{V}_{\mathbf{e}}} \left(\alpha\eta\left(R/\alpha, P, \hat{V}, V\right) + (1-\alpha)\mathsf{D}(U \| W_r | \Pi)\right)} \qquad (24)
\end{aligned}
$$

where $\mathcal{V}_{\mathbf{e}}$ is the complement of $\mathcal{V}_{\mathbf{x}}$ in $\mathcal{V}$ given by

$$
\mathcal{V}_{\mathbf{e}} \triangleq \left\{ (V, \hat{V}, U) : \mathsf{I}(P, V) \geq \mathsf{I}\left(P, \hat{V}\right) \text{ and } (PV)_Y = (P\hat{V})_Y \text{ and } (V, U, m) \succ (\hat{V}, U, m+1) \right\}. \qquad (25)
$$

Note that m in the definition of $\mathcal{V}_{\mathbf{e}}$ is also a dummy variable. The domination rule $\succ$ divides the set $\mathcal{V}$ into two subsets: the erasure subset $\mathcal{V}_{\mathbf{x}}$ and the error subset $\mathcal{V}_{\mathbf{e}}$. Choosing domination rule is equivalent to choosing the $\mathcal{V}_{\mathbf{e}}$. Depending on the value of $\alpha E_r(\frac{R}{\alpha}, P)$ and $E_{\mathbf{x}}$ we chose different $\mathcal{V}_{\mathbf{e}}$'s as follows:

(i) $E_{\mathbf{x}} > \alpha E_r(\frac{R}{\alpha}, P)$: $\mathcal{V}_{\mathbf{e}} = \mathcal{V}$. Then $\mathcal{V}_{\mathbf{x}} = \emptyset$ and Theorem 1 follows from equation (22).

(ii) $E_{\mathbf{x}} \leq \alpha E_r(\frac{R}{\alpha}, P)$: $\mathcal{V}_{\mathbf{e}} = \left\{ (V, \hat{V}, U) : \begin{array}{c} \mathsf{I}(P, V) \geq \mathsf{I}\left(P, \hat{V}\right) \text{ and } (PV)_Y = (P\hat{V})_Y \text{ and} \\ \alpha\eta\left(\frac{R}{\alpha}, P, V, \hat{V}\right) + (1-\alpha)\mathsf{D}(U \| W_a | \Pi) \leq E_{\mathbf{x}} \end{array} \right\}$. Then all the $(V, \hat{V}, U)$ triples satisfying $\alpha\eta\left(\frac{R}{\alpha}, P, V, \hat{V}\right) + (1-\alpha)\mathsf{D}(U \| W_a | \Pi) \leq E_{\mathbf{x}}$ are in the the error subset. Thus as a result of (22) erasure probability is bounded as $P_{\mathbf{x}} \leq e^{-\mathsf{n}(E_{\mathbf{x}} - \delta'_{\mathsf{n}})}$ and Theorem 1 follows from equation (24).

■

E. Lower Bound to $\mathcal{E}_{\mathbf{e}}(R, E_{\mathbf{x}})$:

In this section we use Theorem 1 to derive a lower bound to the optimal error exponent $\mathcal{E}_{\mathbf{e}}(R, E_{\mathbf{x}})$. We do that by optimizing the achievable performance $E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$ over α, P and Π.

1) High Erasure Exponent Region (i.e. $E_{\mathbf{x}} > E_r(R)$): As a result of (18), $\forall R \geq 0$ and $\forall E_{\mathbf{x}} > E_r(R)$

$$
\begin{aligned}
&E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi) = \alpha E_r(\tfrac{R}{\alpha}, P) \leq E_r(R) && \forall \alpha, && \forall P, && \forall \Pi && (26a) \\
&E_{\mathbf{e}}(R, E_{\mathbf{x}}, \tilde{\alpha}, \tilde{P}, \Pi) = E_r(R) && \tilde{\alpha} = 1, && \tilde{P} = \arg\max_{P} E_r(R, P), && \forall \Pi. && (26b)
\end{aligned}
$$

Thus for all $(R, E_{\mathbf{x}})$ pairs such that $E_{\mathbf{x}} > E_r(R)$: optimal time sharing constant is 1, optimal input distribution is the optimal input distribution for random coding exponent at rate R, we use maximum mutual information decoding and never declare erasures. Furthermore since $\alpha = 1$ we have only a single phase in our scheme.

$$
E_{\mathbf{e}}(R, E_{\mathbf{x}}) = E_{\mathbf{e}}(R, E_{\mathbf{x}}, 1, P_{r(R)}, \Pi) = E_r(R) \qquad \forall R \geq 0 \qquad \forall E_{\mathbf{x}} > E_r(R) \qquad (27)
$$

where $P_{r(R)}$ satisfies $E_r(R, P_{r(R)}) = E_r(R)$ and Π can be any control phase type. Evidently benefits of errors-and-erasures decoding is not observed in this region.

2) Low Erasure Exponent Region (i.e. $E_{\mathbf{x}} \leq E_r(R)$): We observe and quantify the benefits of errors-and-erasures decoding for $(R, E_{\mathbf{x}})$ pairs such that $E_{\mathbf{x}} \leq E_r(R)$. Since $E_r(R)$ is a non-negative non-increasing and convex function of R, we have

$$\alpha \in [\alpha^*(R, E_{\mathbf{x}}), 1] \Leftrightarrow E_{\mathbf{x}} \leq \alpha E_r(\tfrac{R}{\alpha}) \qquad \forall R \geq 0 \qquad \forall 0 < E_{\mathbf{x}} \leq E_r(R)$$

where $\alpha^*(R, E_{\mathbf{x}})$ is the unique solution of the equation $\alpha E_r(\frac{R}{\alpha}) = E_{\mathbf{x}}$.

For the case $E_{\mathbf{x}} = 0$, however, $\alpha E_r(\frac{R}{\alpha}) = 0$ has multiple solutions and Theorem 1 holds but resulting error exponent, $E_{\mathbf{e}}(R, 0, \alpha, P, \Pi)$, does not correspond to the error exponent of a reliable sequence. Convention introduced below in equation (28) addresses both issues at once, by choosing the minimum of those solutions as $\alpha^*(R, 0)$. In addition by this convention $\alpha^*(R, E_{\mathbf{x}})$ is also continuous at $E_{\mathbf{x}} = 0$: $\lim_{E_{\mathbf{x}} \to 0} \alpha^*(R, E_{\mathbf{x}}) = \alpha^*(R, 0)$.

$$\alpha^*(R, E_{\mathbf{x}}) \triangleq \begin{cases} \frac{R}{g^{-1}(E_{\mathbf{x}}/R)} & E_{\mathbf{x}} \in (0, E_r(R)] \\ R/\mathcal{C} & E_{\mathbf{x}} = 0 \end{cases} \tag{28}$$

where $g^{-1}(\cdot)$ is the inverse of the function $g(r) = \frac{r}{E_r(r)}$.

As a result equations (18) and (28), $\forall R \geq 0$ and $\forall 0 < E_{\mathbf{x}} \leq E_r(R)$ we have

$$E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi) = \alpha E_r(\tfrac{R}{\alpha}, P) \leq E_r(R) \qquad \forall \alpha \in [0, \alpha^*(R, E_{\mathbf{x}})), \qquad \forall P, \qquad \forall \Pi \tag{29a}$$

$$E_{\mathbf{e}}(R, E_{\mathbf{x}}, \tilde{\alpha}, \tilde{P}, \Pi) = E_r(R) \qquad \tilde{\alpha} = 1, \qquad \tilde{P} = \arg\max_P E_r(R, P), \qquad \forall \Pi. \tag{29b}$$

Thus for all $(R, E_{\mathbf{x}})$ pairs such that $E_{\mathbf{x}} \leq E_r(R)$ optimal time sharing constant is in the interval $[\alpha^*(R, E_{\mathbf{x}}), 1]$.

For an $(R, E_{\mathbf{x}}, \alpha)$ triple such that $R \geq 0$, $E_{\mathbf{x}} \leq E_r(R)$ and $\alpha \in [\alpha^*(R, E_{\mathbf{x}}), 1]$ let $\mathcal{P}(R, E_{\mathbf{x}}, \alpha)$ be

$$\mathcal{P}(R, E_{\mathbf{x}}, \alpha) \triangleq \{P : \alpha E_r(\tfrac{R}{\alpha}, P) \geq E_{\mathbf{x}} \ , \ \mathsf{I}(P, W) \geq \tfrac{R}{\alpha}\}. \tag{30}$$

The constraint on mutual information is there to ensure that $E_{\mathbf{e}}(R, 0, \alpha, P, \Pi)$'s are corresponding to error exponent of reliable sequences. The set $\mathcal{P}(R, E_{\mathbf{x}}, \alpha)$ is convex because $E_r(R, P)$ and $\mathsf{I}(P, W)$ are concave in P.

Note that $\forall R \geq 0$ and $\forall E_{\mathbf{x}} \in (0, E_r(R)]$,

$$E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi) = \alpha E_r(\tfrac{R}{\alpha}, P) \qquad \forall \alpha \in [\alpha^*(R, E_{\mathbf{x}}), 1], \qquad \forall P \notin \mathcal{P}(R, E_{\mathbf{x}}, \alpha), \qquad \forall \Pi \tag{31a}$$

$$E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, \tilde{P}, \Pi) \geq \alpha E_r(\tfrac{R}{\alpha}) \qquad \forall \alpha \in [\alpha^*(R, E_{\mathbf{x}}), 1], \qquad \tilde{P} = \arg\max_P E_r(\tfrac{R}{\alpha}, P), \qquad \forall \Pi. \tag{31b}$$

Thus as a result of (31) we can restrict the optimization over P to $\mathcal{P}(R, E_{\mathbf{x}}, \alpha)$ when $\forall R \geq 0$ and $\forall E_{\mathbf{x}} \in (0, E_r(R)]$. For $E_{\mathbf{x}} = 0$ case if we require the expression $E_{\mathbf{e}}(R, 0, \alpha, P, \Pi)$ to correspond to the error exponent of a reliable sequence, get the restriction given in equation (31). Thus using the definition of $E_{\mathbf{e}}(R, E_{\mathbf{x}})$ given in (42) we get:

$$E_{\mathbf{e}}(R, E_{\mathbf{x}}) = \max_{\alpha \in [\alpha^*(R, E_{\mathbf{x}}), 1]} \max_{P \in \mathcal{P}(R, E_{\mathbf{x}}, \alpha)} \max_{\Pi} E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi) \qquad \forall R \geq 0 \qquad \forall E_{\mathbf{x}} \leq E_r(R) \tag{32}$$

where $\alpha^*(R, E_{\mathbf{x}})$, $\mathcal{P}(R, E_{\mathbf{x}}, \alpha)$ and $E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$ are given in equations (28), (30) and (18).

Unlike $E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$ itself, $E_{\mathbf{e}}(R, E_{\mathbf{x}})$ as defined in (32) corresponds to error exponent of reliable code sequences even at $E_{\mathbf{x}} = 0$.

If maximizing P for the inner maximization in equation (32) is same for all $\alpha \in [\alpha^*(R, E_{\mathbf{x}}), 1]$, the optimal value of α is $\alpha^*(R, E_{\mathbf{x}})$. In order to see that, we first observe that any fixed $(R, E_{\mathbf{x}}, P, \Pi)$ such that $E_r(R, P) \geq E_{\mathbf{x}}$, function $E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$ is convex in α for all $\alpha \in [\alpha^*(R, E_{\mathbf{x}}, P), 1]$ where $\alpha^*(R, E_{\mathbf{x}}, P)$ is the unique solution of the equation[12] $\alpha E_r(\frac{R}{\alpha}, P) = E_{\mathbf{x}}$ as it is shown Lemma 10 in Appendix B. Since the maximization preserves the convexity, $\max_{\Pi} E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$ is also convex in α for all $\alpha \in [\alpha^*(R, E_{\mathbf{x}}, P), 1]$. Thus for any $(R, E_{\mathbf{x}}, P)$ triple, $\max_{\Pi} E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$, takes its maximum value either at the minimum possible value of α, i.e. $\alpha^*(R, E_{\mathbf{x}}, P) = \alpha^*(R, E_{\mathbf{x}})$, or at the maximum possible value of α, i.e. 1. It is shown in Appendix C $\max_{\Pi} E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$ takes its is maximum value at $\alpha = \alpha^*(R, E_{\mathbf{x}})$.

[12]Evidently we need to make a minor modification for $E_{\mathbf{x}} = 0$ case as before to ensure that we consider only the $\tilde{E}_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$'s that correspond to the reliable sequences: $\alpha^*(R, 0, P) = \frac{R}{\mathsf{I}(P,W)}$.

Furthermore if the maximizing P is not only the same for all $\alpha \in [\alpha^*(R,E_{\mathbf{x}}),1]$ for a given $(R,E_{\mathbf{x}})$ pair but also for all $(R,E_{\mathbf{x}})$ pairs such that $E_{\mathbf{x}} \leq E_r(R)$ then we can find the optimal $E_{\mathbf{e}}(R,E_{\mathbf{x}})$ by simply maximizing over Π's. In symmetric channels, for example, uniform distribution is the optimal distribution for all $(R,E_{\mathbf{x}})$ pairs. Thus

$$E_{\mathbf{e}}(R,E_{\mathbf{x}}) = \left\{ \begin{array}{cl} E_{\mathbf{e}}(R,E_{\mathbf{x}},1,P^*,\Pi) & \text{if } E_{\mathbf{x}} > E_r(R,P^*) \\ \max_{\Pi} E_{\mathbf{e}}(R,E_{\mathbf{x}},\alpha^*(R,E_{\mathbf{x}}),P^*,\Pi) & \text{if } E_{\mathbf{x}} \leq E_r(R,P^*) \end{array} \right\} \tag{33}$$

where P^* is the uniform distribution.

F. Alternative Expression for Exponent:

The minimization given in (18) for $E_{\mathbf{e}}(R,E_{\mathbf{x}},\alpha,P,\Pi)$ is over transition probability matrices and control phase output types. In order to get a better grasp of the resulting expression, we simplify the analytical expression in this section. We do that by expressing the minimization in (18) in terms of solutions of lower dimensional optimization problems.

Let $\zeta(R,P,Q)$ be the minimum Kullback-Leibler divergence under P with respect to W among the transition probability matrices whose mutual information under P is less than R and whose output distribution under P is Q. It is shown in Appendix B that for a given P, $\zeta(R,P,Q)$ is convex in (R,Q) pair. Evidently for a given (P,Q) pair $\zeta(R,P,Q)$ is a non-increasing in R. Thus for a given (P,Q) pair $\zeta(R,P,Q)$ is strictly decreasing on a closed interval and is an extended real valued function of the form:

$$\zeta(R,P,Q) = \left\{ \begin{array}{cl} \infty & R < R_l^*(P,Q) \\ \min_{\substack{V:\mathsf{I}(P,V)\leq R \\ (PV)_Y=Q}} \mathsf{D}\left(V\|\,W|P\right) & R \in [R_l^*(P,Q),R_h^*(P,Q)] \\ \min_{V:(PV)_Y=Q} \mathsf{D}\left(V\|\,W|P\right) & R > R_h^*(P,Q) \end{array} \right\} \tag{34a}$$

$$R_l^*(P,Q) = \min_{\substack{V:\,PV\gg PW \\ (PV)_Y=Q}} \mathsf{I}\left(P,V\right) \tag{34b}$$

$$R_h^*(P,Q) = \min_R \left\{ R : \min_{\substack{V:\mathsf{I}(P,V)\leq R \\ (PV)_Y=Q}} \mathsf{D}\left(V\|\,W|P\right) = \min_{V:(PV)_Y=Q} \mathsf{D}\left(V\|\,W|P\right) \right\} \tag{34c}$$

where $PV \gg PW$ iff for all (x,y) pairs such that $P(x)W(y|x)$ is zero, $P(x)V(y|x)$ is also zero.

Let $\Gamma\left(T,\Pi\right)$ be the minimum Kullback-Leibler divergence with respect to W_r under Π, among the U's whose Kullback-Leibler divergence with respect to W_a under Π is less than or equal to T.

$$\Gamma\left(T,\Pi\right) \triangleq \min_{U:\mathsf{D}\left(U\|\,W_a|\Pi\right)\leq T} \mathsf{D}\left(U\|\,W_r|\Pi\right) \tag{35}$$

For a given Π, $\Gamma\left(T,\Pi\right)$ is non-increasing and convex in T, thus $\Gamma\left(T,\Pi\right)$ is strictly decreasing in T on a closed interval. An equivalent expressions for $\Gamma\left(T,\Pi\right)$ and boundaries of this closed interval is derived in Appendix A,

$$\Gamma\left(T,\Pi\right) = \left\{ \begin{array}{ccll} \infty & \text{if} & T < \mathsf{D}\left(U_0\|\,W_a|\Pi\right) & \\ \mathsf{D}\left(U_s\|\,W_r|\Pi\right) & \text{if} & T = \mathsf{D}\left(U_s\|\,W_a|\Pi\right) & \text{for some } s\in[0,1] \\ \mathsf{D}\left(U_1\|\,W_r|\Pi\right) & \text{if} & T > \mathsf{D}\left(U_1\|\,W_a|\Pi\right) & \end{array} \right\} \tag{36}$$

where

$$U_s(y|x_1,x_2) = \left\{ \begin{array}{lll} \frac{\mathbb{1}_{\{W(y|x_2)>0\}}}{\sum_{\tilde{y}:W(\tilde{y}|x_2)>0} W(\tilde{y}|x_1)} W(y|x_1) & \text{if} & s=0 \\ \frac{W(y|x_1)^{1-s}W(y|x_2)^s}{\sum_{\tilde{y}} W(\tilde{y}|x_1)^{1-s}W(\tilde{y}|x_2)^s} & \text{if} & s\in(0,1) \\ \frac{\mathbb{1}_{\{W(y|x_1)>0\}}}{\sum_{\tilde{y}:W(\tilde{y}|x_1)>0} W(\tilde{y}|x_2)} W(y|x_2) & \text{if} & s=1 \end{array} \right\}$$

For a $(R,E_{\mathbf{x}},\alpha,P,\Pi)$ such that $E_{\mathbf{x}} \leq \alpha E_r(\frac{R}{\alpha},P)$, using the definition of $E_{\mathbf{e}}(R,E_{\mathbf{x}},\alpha,P,\Pi)$ in (18) together with the equations (13), (34) and (36) we get

$$E_{\mathbf{e}}(R,E_{\mathbf{x}},\alpha,P,\Pi) = \min_{\substack{Q,T,R_1,R_2: \\ R_1\geq R_2\geq 0,\ T\geq 0 \\ \alpha\zeta(\frac{R_1}{\alpha},P,Q)+|R_2-R|^+ +T\leq E_{\mathbf{x}}}} \alpha\zeta(\tfrac{R_2}{\alpha},P,Q)+|R_1-R|^+ +(1-\alpha)\Gamma\left(\tfrac{T}{1-\alpha},\Pi\right)$$

For any $(R,E_{\mathbf{x}},\alpha,P,\Pi)$ above minimum is also achieved at a (Q,R_1,R_2,T) such that $R_1 \geq R_2 \geq R$. In order to see this take any minimizing (Q^*,R_1^*,R_2^*,T^*), then there are three possibilities:

(a) $R_1^* \geq R_2^* \geq R$ claim holds trivially.
(b) $R_1^* \geq R > R_2^*$, since $\zeta(\frac{R_2}{\alpha}, P, Q)$ is non-increasing function (Q^*, R_1^*, R, T^*), is also minimizing, thus claim holds.
(c) $R > R_1^* > R_2^*$, since $\zeta(\frac{R}{\alpha}, P, Q)$ is non-increasing function (Q^*, R, R, T^*), is also minimizing, thus claim holds.
Thus we obtain the following expression for $E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$,

$$E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi) = \left\{ \begin{array}{ll} \alpha E_r(\frac{R}{\alpha}, P) & \text{if } E_{\mathbf{x}} > \alpha E_r(\frac{R}{\alpha}, P) \\ \min\limits_{\substack{Q,T,R_1,R_2: \\ R_1 \geq R_2 \geq R,\ T \geq 0 \\ \alpha\zeta(\frac{R_1}{\alpha}, P, Q) + R_2 - R + T \leq E_{\mathbf{x}}}} \alpha\zeta(\frac{R_2}{\alpha}, P, Q) + R_1 - R + (1-\alpha)\Gamma\left(\frac{T}{1-\alpha}, \Pi\right) & \text{if } E_{\mathbf{x}} \leq \alpha E_r(\frac{R}{\alpha}, P) \end{array} \right\} \tag{37}$$

Equation (37) is simplified further for symmetric channels. For symmetric channels,

$$E_{sp}(R) = \zeta(R, P^*, Q^*) = \min_{Q} \zeta(R, P^*, Q) \tag{38}$$

where P^* is the uniform input distribution and Q^* is the corresponding output distribution under W.

Using alternative expression for $E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$ given in (37) together with equations (33) and (38) for symmetric channels we get,

$$E_{\mathbf{e}}(R, E_{\mathbf{x}}) = \left\{ \begin{array}{ll} E_r(R) & \text{if } E_{\mathbf{x}} > E_r(R) \\ \max\limits_{\Pi} \min\limits_{\substack{R'', R', T: \\ R'' \geq R' \geq R\ T \geq 0 \\ \alpha^* E_{sp}(\frac{R''}{\alpha^*}) + R' - R + T \leq E_{\mathbf{x}}}} \alpha^* E_{sp}\left(\frac{R'}{\alpha^*}\right) + R'' - R + (1-\alpha^*)\Gamma\left(\frac{T}{1-\alpha^*}, \Pi\right) & \text{if } E_{\mathbf{x}} \leq E_r(R) \end{array} \right\} \tag{39}$$

where $\alpha^*(R, E_{\mathbf{x}})$ is given in equation (28).

Although (38) does not hold in general using definition of $\zeta(R, P, Q)$ and $E_{sp}(R, P)$ we can assert that

$$\zeta(R, P, Q) \geq \min_{\tilde{Q}} \zeta(R, P, \tilde{Q}) = E_{sp}(R, P) \tag{40}$$

Note that (40) can be used to bound the minimized expression in (37) from below. In addition recall that if the set that a minimization is done over is enlarged resulting minimum can not increase. We can use (37) also to enlarge the set that minimization is done over in (40). Thus we get an exponent $\tilde{E}_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$ which is smaller than or equal to $E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$ in all channels and for all $\tilde{E}_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$'s:

$$\tilde{E}_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi) = \left\{ \begin{array}{ll} \alpha E_r(\frac{R}{\alpha}, P) & \text{if } E_{\mathbf{x}} > \alpha E_r(\frac{R}{\alpha}, P) \\ \min\limits_{\substack{R'', R', T: \\ R'' \geq R' \geq R\ T \geq 0 \\ \alpha E_{sp}(\frac{R''}{\alpha}, P) + R' - R + T \leq E_{\mathbf{x}}}} \alpha E_{sp}\left(\frac{R'}{\alpha}, P\right) + R'' - R + (1-\alpha)\Gamma\left(\frac{T}{1-\alpha}, \Pi\right) & \text{if } E_{\mathbf{x}} \leq \alpha E_r(\frac{R}{\alpha}, P) \end{array} \right\} \tag{41}$$

After an investigation very similar to the one we have already done for $E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$ in Section III-E, we obtain the below expression for the optimal error exponent for reliable sequences emerging from (41):

$$\tilde{E}_{\mathbf{e}}(R, E_{\mathbf{x}}) = \left\{ \begin{array}{ll} E_r(R) & \forall R \geq 0 \quad \forall E_{\mathbf{x}} > E_r(R) \\ \max\limits_{\alpha \in [\alpha^*(R, E_{\mathbf{x}}), 1]} \max\limits_{P \in \mathcal{P}(R, E_{\mathbf{x}}, \alpha)} \max\limits_{\Pi} \tilde{E}_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi) & \forall R \geq 0 \quad \forall E_{\mathbf{x}} \leq E_r(R) \end{array} \right\} \tag{42}$$

where $\alpha^*(R, E_{\mathbf{x}})$, $\mathcal{P}(R, E_{\mathbf{x}}, \alpha)$ and $\tilde{E}_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$ are given in equations (28), (30) and (41), respectively.

G. Special Cases

1) Zero Erasure Exponent Case, $\mathcal{E}_{\mathbf{e}}(R, 0)$: Using a simple repetition-at-erasures scheme, fixed length errors and erasures codes, can be converted into variable length block-codes, with the same error exponent. Thus the error exponents of variable length block-codes given by Burnashev in [3] is an upper bound to the error exponent of fixed length block-codes with erasures:

$$\mathcal{E}_{\mathbf{e}}(R, E_{\mathbf{x}}) \leq \left(1 - \tfrac{R}{\mathcal{C}}\right)\mathcal{D} \qquad \forall R \geq 0, E_{\mathbf{x}} \geq 0$$

where $\mathcal{D} = \max_{x, \tilde{x}} \sum_y W(y|x) \log \frac{W(y|x)}{W(y|\tilde{x})}$.

We show below that, $\tilde{E}_{\mathbf{e}}(R,0) \geq (1-\frac{R}{\mathcal{C}})\mathcal{D}$. This implies that our coding scheme is optimal for $E_{\mathbf{x}}=0$ for all rates i.e. $\tilde{E}_{\mathbf{e}}(R,0) = \mathcal{E}_{\mathbf{e}}(R,0) = \mathcal{D}$.

Recall that for all R less than capacity $\alpha^*(R,0)=\frac{R}{\mathcal{C}}$. Furthermore for any $\alpha \geq \frac{R}{\mathcal{C}}$

$$\mathcal{P}\left(R,0,\alpha\right) = \{P : \mathsf{I}\left(P,W\right) \geq \tfrac{R}{\alpha}\}$$

Thus for any $(R,0,\alpha,P)$ such that $P \in \mathcal{P}\left(R,0,\alpha\right)$, $R'' \geq R' \geq R$, $T \geq 0$ and $\alpha E_{sp}(\frac{R''}{\alpha},P) + R' - R + T \leq 0$, imply that $R'=R$, $R''=\alpha\mathsf{I}\left(P,W\right)$, $T=0$. Consequently

$$\tilde{E}_{\mathbf{e}}(R,0,\alpha,P,\Pi) = \alpha\left[E_{sp}\left(\tfrac{R}{\alpha},P\right) + \mathsf{I}\left(P,W\right) - \tfrac{R}{\alpha}\right] + (1-\alpha)\mathsf{D}\left(W_r\| W_a|\Pi\right) \tag{43}$$

When we maximize over Π and $P \in \mathcal{P}\left(R,0,\alpha\right)$ we get:

$$\tilde{E}_{\mathbf{e}}(R,0,\alpha) = \max_{P\in\mathcal{P}(R,0,\alpha)} \alpha E_{sp}\left(\tfrac{R}{\alpha},P\right) + \alpha\mathsf{I}\left(P,W\right) - R + (1-\alpha)\mathcal{D} \qquad \forall\alpha\in[\tfrac{R}{\mathcal{C}},1]. \tag{44}$$

If simply insert the minimum possible value of α i.e. $\alpha^*(R,0)=\frac{R}{\mathcal{C}}$:

$$\begin{aligned}\tilde{E}_{\mathbf{e}}(R,0,\tfrac{R}{\mathcal{C}}) &= \max_{P\in\mathcal{P}\left(R,0,\frac{R}{\mathcal{C}}\right)} \tfrac{R}{\mathcal{C}}E_{sp}\left(\mathcal{C},P\right) + \tfrac{R}{\mathcal{C}}\mathsf{I}\left(P,W\right) - R + (1-\tfrac{R}{\mathcal{C}})\mathcal{D}\\ &= (1-\tfrac{R}{\mathcal{C}})\mathcal{D}.\end{aligned}$$

Thus $\tilde{E}_{\mathbf{e}}(R,0) \geq (1-\frac{R}{\mathcal{C}})\mathcal{D}$.

Indeed one need not to rely on the converse on variable length block-codes in order to establish the fact that $\tilde{E}_{\mathbf{e}}(R,0) = (1-\frac{R}{\mathcal{C}})\mathcal{D}$. The lower bound to probability of error presented in the next section, not only recovers this particular optimality result but also upper bounds the optimal error exponent, $\mathcal{E}_{\mathbf{e}}(R,E_{\mathbf{x}})$, as a function of rate R and erasure exponents $E_{\mathbf{x}}$.

2) Channels with non-zero Zero Error Capacity: For channels with a non-zero zero-error capacity, one can use equation (18) to prove that, for any $E_{\mathbf{x}} < E_r(R)$, $E_{\mathbf{e}}(R,E_{\mathbf{x}}) = \infty$. This implies that we can get error-free block-codes with this two phase coding scheme for any rate $R < C$ and any erasure exponent $E_{\mathbf{x}} \leq E_r(R)$. As we discuss in Section V in more detail, this is the best erasure exponent for rates over the critical rate.

IV. An Outer Bound for Error Exponent Erasure Exponent Trade-off

In this section we derive an upper bound on $\mathcal{E}_{\mathbf{e}}(R,E_{\mathbf{x}})$ using previously known results on erasure free block-codes with feedback and a generalization of the straight line bound of Shannon, Gallager and Berlekamp [29]. We first present a lower bound on the minimum error probability of block-codes with feedback and erasures, in terms of that of shorter codes in Section IV-A. Then in Section IV-B we make a brief overview of the outer bounds on the error exponents of erasure free block-codes with feedback. Finally in Section IV-C, we use the relation we have derived in Section IV-A to tie the previously known results we have summarized in Section IV-B to bound $\mathcal{E}_{\mathbf{e}}(R,E_{\mathbf{x}})$.

A. A Trait of Minimum Error Probability of block-codes with Erasures

Shannon, Gallager and Berlekamp in [29] considered fixed length block-codes, with list decoding and established a family of lower bounds on the minimum error probability in terms of the product of minimum error probabilities of certain shorter codes. They have shown, [29, Theorem 1], that for fixed length block-codes with list decoding and without feedback

$$\tilde{\mathcal{P}_{\mathbf{e}}}(M,\mathsf{n},L) \geq \tilde{\mathcal{P}_{\mathbf{e}}}(M,\mathsf{n}_1,L_1)\tilde{\mathcal{P}_{\mathbf{e}}}(L_1+1,\mathsf{n}-\mathsf{n}_1,L) \tag{45}$$

where $\tilde{\mathcal{P}_{\mathbf{e}}}(M,\mathsf{n},L)$ denotes the minimum error probability of erasure free block codes of length n with M equally probable messages and with decoding list size L. As they have already pointed out in [29] this theorem continues to hold in the case when a feedback link is available from receiver to the transmitter; although $\tilde{\mathcal{P}_{\mathbf{e}}}$'s are different when feedback is available, the relation given in equation (45) still holds. They were interested in erasure free codes. We, on the other hand, are interested in block-codes which might have non-zero erasure probability. Accordingly we need to incorporate erasure probability as one of the parameters of the optimal error probability. This is what this section is dedicated to.

Decoded set $\hat{\mathsf{M}}$ of a size L list decoder with erasures is either a subset[13] of $\mathcal{M}$ whose size is at most L, like the erasure-free case, or a set which only includes the erasure symbol, i.e. either $\hat{\mathsf{M}} \subset \mathcal{M}$ such that $|\hat{\mathsf{M}}| \leq L$ or $\hat{\mathsf{M}} = \{\mathbf{x}\}$. The minimum error probability of length n block-codes, with M equally probable messages, decoding list size L and erasure probability $P_{\mathbf{x}}$ is given by $\mathcal{P}_{\mathbf{e}}(M, \mathsf{n}, L, P_{\mathbf{x}})$.

Theorem 2 below bounds the error probability of block codes with erasures and list decoding using the error probabilities of shorter codes with erasures and list decoding, like [29, Theorem 1] does in the erasure free case. Like its counter part in erasure free case Theorem 2 is later used to establish outer bounds to error exponents.

Theorem 2: For any n, M, L, $P_{\mathbf{x}}$, $\mathsf{n}_1 \leq \mathsf{n}$, L_1, and $0 \leq s \leq 1$ the minimum error probability of fixed length block-codes with feedback satisfy

$$\mathcal{P}_{\mathbf{e}}(M, \mathsf{n}, L, P_{\mathbf{x}}) \geq \mathcal{P}_{\mathbf{e}}(M, \mathsf{n}_1, L_1, s)\mathcal{P}_{\mathbf{e}}\left(L_1 + 1, \mathsf{n} - \mathsf{n}_1, L, \tfrac{(1-s)P_{\mathbf{x}}}{\mathcal{P}_{\mathbf{e}}(M, \mathsf{n}_1, L_1, s)}\right) \tag{46}$$

Let us first consider the following lemma which bounds the achievable error probability-erasure probability pairs for block-codes with nonuniform a priori probability distribution, in terms of block codes with a uniform a priori probability distribution but fewer messages.

Lemma 2: For any length n block-code with message set $\mathcal{M}$, a priori probability distribution $\varphi(\cdot)$ on $\mathcal{M}$, erasure probability $P_{\mathbf{x}}$, list decoding size L, and any integer K

$$P_{\mathbf{e}} \geq \Omega(\varphi, K)\, \mathcal{P}_{\mathbf{e}}\left(K + 1, \mathsf{n}, L, \tfrac{P_{\mathbf{x}}}{\Omega(\varphi, K)}\right) \qquad \text{where} \qquad \Omega(\varphi, K) = \min_{\mathcal{S} \subset \mathcal{M}: |\mathcal{S}| = |\mathcal{M}| - K} \varphi(\mathcal{S}). \tag{47}$$

where $\mathcal{P}_{\mathbf{e}}(K + 1, \mathsf{n}, L, P_{\mathbf{x}})$ is the minimum error probability of length n codes with $(K+1)$ equally probable messages and decoding list size L, with feedback if the original code has feedback without feedback if the original code has not.

Note that $\Omega(\varphi, K)$ is the minimum error probability of a size K decoder, if the posterior probability distribution on the messages is φ.

Proof: If $\Omega(\varphi, K) = 0$ theorem holds trivially. Thus we assume $\Omega(\varphi, K) > 0$ henceforth. For any size $(K + 1)$ subset $\mathcal{M}'$ of $\mathcal{M}$, we can use the encoding scheme and the decoding rule of the original code for $\mathcal{M}$, to construct the following block-code for $\mathcal{M}'$. For each $m \in \mathcal{M}'$ use the encoding scheme for message m in the original code, i.e.

$$X'_t(m, y^{t-1}) = X_t(m, y^{t-1}) \qquad \forall m \in \mathcal{M}', \quad t \in [1, \mathsf{n}], \quad y^{t-1} \in \mathcal{Y}^{t-1}$$

For all $y^{\mathsf{n}} \in \mathcal{Y}^{\mathsf{n}}$, if the original decoding rule declares erasure, decoding rule of the new code declares erasure, else the decoded list is the intersection of the original decoded list with $\mathcal{M}'$.

$$\hat{\mathsf{M}}' = \begin{cases} \mathbf{x} & \text{if } \hat{\mathsf{M}} = \mathbf{x} \\ \hat{\mathsf{M}} \cap \mathcal{M}' & \text{else} \end{cases}$$

Note that this is a length n code with $(K + 1)$ messages and list decoding size L. Furthermore for all $m \in \mathcal{M}'$ the conditional error and erasure probabilities, ${P_{\mathbf{x}}}'(m)$, ${P_{\mathbf{e}}}'(m)$ are equal to the conditional error and erasure probabilities in the original code, $P_{\mathbf{x}|m}$, $P_{\mathbf{e}|m}$. Thus

$$\tfrac{1}{K+1} \sum_{m \in \mathcal{M}'} \left(P_{\mathbf{x}|m}, P_{\mathbf{e}|m}\right) \in \Psi(K + 1, \mathsf{n}, L) \qquad \forall \mathcal{M}' \subset \mathcal{M} \text{ such that } |\mathcal{M}'| = K + 1 \tag{48}$$

where $\Psi(K + 1, \mathsf{n}, L)$ is the set of achievable error probability, erasure probability pairs for length n block-codes with $(K + 1)$ equally probable messages and with decoding list size L. Evidently $\Psi(M, \mathsf{n}, L)$ is a convex set for any (M, n, L) triple. Furthermore, $\forall a \geq 1$, $\forall b_1 \geq 0$, $\forall b_2 \geq 0$:

$$\psi \in \Psi \Rightarrow (a \cdot \psi + (b_1, b_2)) \in \Psi. \tag{49}$$

Note that $\Psi(M, \mathsf{n}, L)$ is uniquely determined by $\mathcal{P}_{\mathbf{e}}(M, \mathsf{n}, L, s_{\mathbf{x}})$ for $s_{\mathbf{x}} \in [0, 1]$ and vice verse:

$$\mathcal{P}_{\mathbf{e}}(M, \mathsf{n}, L, \psi_{\mathbf{x}}) = \min_{\psi_{\mathbf{x}}: (\psi_{\mathbf{e}}, \psi_{\mathbf{x}}) \in \Psi(M, \mathsf{n}, L)} \psi_{\mathbf{e}} \qquad \forall (M, \mathsf{n}, L, \psi_{\mathbf{x}}). \tag{50}$$

[13]Note that if $\hat{\mathsf{M}} \subset \mathcal{M}$ then $\mathbf{x} \notin \hat{\mathsf{M}}$ because $\mathbf{x} \notin \mathcal{M}$.

Let the smallest non-zero element of $\{\varphi(1), \varphi(2), \dots \varphi(|\mathcal{M}|)\}$ be $\varphi(\xi_1)$. For any size $(K+1)$ subset of $\mathcal{M}$ which includes ξ_1 and all whose elements have non-zero probabilities, say $\mathcal{M}_1$, we have,

$$\begin{aligned}(P_{\mathbf{x}}, P_{\mathbf{e}}) &= \sum\nolimits_{m\in\mathcal{M}} \varphi(m)(P_{\mathbf{x}|m}, P_{\mathbf{e}|m}) \\ &= \sum\nolimits_{m\in\mathcal{M}} [\varphi(m) - \varphi(\xi_1)\mathbb{1}_{\{m\in\mathcal{M}_1\}}](P_{\mathbf{x}|m}, P_{\mathbf{e}|m}) + \varphi(\xi_1)\sum\nolimits_{m\in\mathcal{M}_1}(P_{\mathbf{x}|m}, P_{\mathbf{e}|m})\end{aligned}$$

As result of equation (48) and the definition of $\Psi(K+1, \mathsf{n}, L)$ we can conclude that $\exists\psi_1 \in \Psi(K+1, \mathsf{n}, L)$ such that

$$(P_{\mathbf{x}}, P_{\mathbf{e}}) = \sum\nolimits_{m\in\mathcal{M}} \varphi^{(1)}(m)(P_{\mathbf{x}|m}, P_{\mathbf{e}|m}) + \varphi(\psi_1)\psi_1 \tag{51}$$

where $\varphi(\psi_1) = (K+1)\varphi(\xi_1)$ and $\varphi^{(1)}(m) = \varphi(m) - \varphi(\xi_1)\mathbb{1}_{\{m\in\mathcal{M}_1\}}$. Consequently

$$\varphi(\psi_1) + \sum\nolimits_{m\in\mathcal{M}} \varphi^{(1)}(m) = 1 \tag{52}$$

Furthermore the number of non-zero $\varphi^{(1)}(m)$'s is at least one less than that of non-zero $\varphi(m)$'s. The remaining probabilities, $\varphi^{(1)}(m)$, have a minimum, $\varphi^{(1)}(\xi_2)$ among its non-zero elements. We can repeat the same argument once more using that element and reduce the number of non-zero elements at least one more. After at most $|\mathcal{M}| - K$ such iterations we reach to a $\varphi^{(\ell)}$ which is non-zero for K or fewer messages:

$$(P_{\mathbf{x}}, P_{\mathbf{e}}) = \sum\nolimits_{j=1}^{\ell} \varphi^{(j-1)}(\psi_j)\psi_j + \sum\nolimits_{m\in\mathcal{M}} \varphi^{(\ell)}(m)(P_{\mathbf{x}|m}, P_{\mathbf{e}|m}) \tag{53}$$

where $\varphi^{(\ell)}(m) \le \varphi(m)$ for all m in $\mathcal{M}$ and $\sum_{m\in\mathcal{M}} \mathbb{1}_{\{\varphi^{(\ell)}(m)>0\}} \le K$. Thus as a result of definition of $\Omega(\varphi, K)$ given in equation (47),

$$\Omega(\varphi, K) \le \sum\nolimits_{j=1}^{\ell} \varphi^{(j-1)}(\psi_j). \tag{54}$$

Note that in equation (53), the first sum is equal to a convex combination of ψ_j's multiplied by $\sum_{j=1}^{\ell} \varphi(\psi_j)$; the second sum is equal to a pair with non-negative entries. Using the convexity of $\Psi(K+1, \mathsf{n}, L)$, the identity in (49) and the equation (54) we see that

$$\exists\psi \in \Psi(K+1, \mathsf{n}, L) \text{ such that } (P_{\mathbf{x}}, P_{\mathbf{e}}) = \Omega(\varphi, K)\,\psi \tag{55}$$

The lemma follows equations (55) and (50). ■

For proving Theorem 2, we express the error and erasure probabilities, as a convex combination of error and erasure probabilities of $(\mathsf{n}-\mathsf{n}_1)$ long block-codes with a priori probability distribution $\varphi_{y^{\mathsf{n}_1}}(m) = \mathbf{P}[m|\, y^{\mathsf{n}_1}]$ over the messages and apply Lemma 2 together with convexity arguments similar to the ones above.

Proof [Theorem 2]:
For all m in $\mathcal{M}$, let $\Upsilon(m)$ be the decoding region of m, $\Upsilon(\mathbf{x})$ be the decoding region of the erasure symbol $\mathbf{x}$ and $\tilde{\Upsilon}(m)$ the error region of m:

$$\Upsilon(m) = \{y^{\mathsf{n}} : m \in \hat{\mathsf{M}}\} \qquad \Upsilon(\mathbf{x}) = \{y^{\mathsf{n}} : \mathbf{x} \in \hat{\mathsf{M}}\} \qquad \tilde{\Upsilon}(m) = \Upsilon(m)^c \cap \Upsilon(\mathbf{x})^c \tag{56}$$

Then for all $m \in \mathcal{M}$ we have

$$(P_{\mathbf{x}|m}, P_{\mathbf{e}|m}) = \left(\mathbf{P}[\Upsilon(\mathbf{x})|\, m]\,, \mathbf{P}\left[\tilde{\Upsilon}(m)\middle|\, m\right]\right) \tag{57}$$

Note that

$$\begin{aligned}P_{\mathbf{x}|m} &= \sum\nolimits_{y^{\mathsf{n}}:y^{\mathsf{n}}\in\Upsilon(\mathbf{x})} \mathbf{P}[y^{\mathsf{n}}|\, m] \\ &= \sum\nolimits_{y^{\mathsf{n}_1}} \mathbf{P}[y^{\mathsf{n}_1}|\, m] \sum\nolimits_{y^{\mathsf{n}}_{\mathsf{n}_1+1}:(y^{\mathsf{n}_1}, y^{\mathsf{n}}_{\mathsf{n}_1+1})\in\Upsilon(\mathbf{x})} \mathbf{P}\left[y^{\mathsf{n}}_{\mathsf{n}_1+1}\middle|\, m, y^{\mathsf{n}_1}\right]\end{aligned}$$

Then the erasure probability is

$$\begin{aligned}P_{\mathbf{x}} &= \sum\nolimits_{m\in\mathcal{M}} \tfrac{1}{|\mathcal{M}|} \sum\nolimits_{y^{\mathsf{n}_1}} \mathbf{P}[y^{\mathsf{n}_1}|\, m] \sum\nolimits_{y^{\mathsf{n}}_{\mathsf{n}_1+1}:(y^{\mathsf{n}_1}, y^{\mathsf{n}}_{\mathsf{n}_1+1})\in\Upsilon(\mathbf{x})} \mathbf{P}\left[y^{\mathsf{n}}_{\mathsf{n}_1+1}\middle|\, m, y^{\mathsf{n}_1}\right] \\ &= \sum\nolimits_{y^{\mathsf{n}_1}} \mathbf{P}[y^{\mathsf{n}_1}] \left(\sum\nolimits_{m\in\mathcal{M}} \mathbf{P}[m|\, y^{\mathsf{n}_1}] \sum\nolimits_{y^{\mathsf{n}}_{\mathsf{n}_1+1}:(y^{\mathsf{n}_1}, y^{\mathsf{n}}_{\mathsf{n}_1+1})\in\Upsilon(\mathbf{x})} \mathbf{P}\left[y^{\mathsf{n}}_{\mathsf{n}_1+1}\middle|\, m, y^{\mathsf{n}_1}\right]\right) \\ &= \sum\nolimits_{y^{\mathsf{n}_1}} \mathbf{P}[y^{\mathsf{n}_1}]\, P_{\mathbf{x}}(y^{\mathsf{n}_1})\end{aligned}$$

Note that for every y^{n_1}, $P_{\mathbf{x}}(y^{\mathsf{n}_1})$ is the erasure probability of a code of length $(\mathsf{n}-\mathsf{n}_1)$ with a priori probability distribution is $\varphi_{y^{\mathsf{n}_1}}(m) = \mathbf{P}[m|\, y^{\mathsf{n}_1}]$. Furthermore we can write the error probability, $P_{\mathbf{e}}$ as

$$P_{\mathbf{e}} = \sum\nolimits_{y^{\mathsf{n}_1}} \mathbf{P}[y^{\mathsf{n}_1}] \left(\sum_{m\in\mathcal{M}} \mathbf{P}[m|\, y^{\mathsf{n}_1}] \sum\nolimits_{y^{\mathsf{n}}_{\mathsf{n}_1+1}:(y^{\mathsf{n}_1}, y^{\mathsf{n}}_{\mathsf{n}_1+1})\in\tilde{\Upsilon}(m)} \mathbf{P}\left[y^{\mathsf{n}}_{\mathsf{n}_1+1}\middle|\, m, y^{\mathsf{n}_1}\right] \right)$$
$$= \sum\nolimits_{y^{\mathsf{n}_1}} \mathbf{P}[y^{\mathsf{n}_1}]\, P_{\mathbf{e}}(y^{\mathsf{n}_1})$$

where $P_{\mathbf{e}}(y^{\mathsf{n}_1})$ is the error probability of the very same length $(\mathsf{n}-\mathsf{n}_1)$ code. As a result of Lemma 2 we know that the pair $(P_{\mathbf{e}}(y^{\mathsf{n}_1}), P_{\mathbf{x}}(y^{\mathsf{n}_1}))$ satisfies

$$P_{\mathbf{e}}(y^{\mathsf{n}_1}) \geq \Omega\left(\varphi_{y^{\mathsf{n}_1}}, L_1\right) \mathcal{P}_{\mathbf{e}}\left(L_1+1, (\mathsf{n}-\mathsf{n}_1), L, \tfrac{P_{\mathbf{x}}(y^{\mathsf{n}_1})}{\Omega\left(\varphi_{y^{\mathsf{n}_1}}, L_1\right)}\right) \tag{58}$$

Then for any $s \in [0,1]$.

$$\begin{aligned} P_{\mathbf{e}} &\geq \sum\nolimits_{y^{\mathsf{n}_1}} \mathbf{P}[y^{\mathsf{n}_1}]\,(1-s)\Omega\left(\varphi_{y^{\mathsf{n}_1}}, L_1\right) \mathcal{P}_{\mathbf{e}}\left(L_1+1, (\mathsf{n}-\mathsf{n}_1), L, \tfrac{P_{\mathbf{x}}(y^{\mathsf{n}_1})}{\Omega\left(\varphi_{y^{\mathsf{n}_1}}, L_1\right)}\right) \\ &\geq \left(\sum\nolimits_{y^{\mathsf{n}_1}} \mathbf{P}[y^{\mathsf{n}_1}]\,(1-s)\Omega\left(\varphi_{y^{\mathsf{n}_1}}, L_1\right)\right) \mathcal{P}_{\mathbf{e}}\left(L_1+1, (\mathsf{n}-\mathsf{n}_1), L, \tfrac{\sum_{y^{\mathsf{n}_1}} \mathbf{P}[y^{\mathsf{n}_1}](1-s)P_{\mathbf{x}}(y^{\mathsf{n}_1})}{\sum_{y^{\mathsf{n}_1}} \mathbf{P}[y^{\mathsf{n}_1}](1-s)\Omega\left(\varphi_{y^{\mathsf{n}_1}}, L_1\right)}\right) \\ &= \left(\sum\nolimits_{y^{\mathsf{n}_1}} \mathbf{P}[y^{\mathsf{n}_1}]\,(1-s)\Omega\left(\varphi_{y^{\mathsf{n}_1}}, L_1\right)\right) \mathcal{P}_{\mathbf{e}}\left(L_1+1, (\mathsf{n}-\mathsf{n}_1), L, \tfrac{(1-s)P_{\mathbf{x}}}{\sum_{y^{\mathsf{n}_1}} \mathbf{P}[y^{\mathsf{n}_1}](1-s)\Omega\left(\varphi_{y^{\mathsf{n}_1}}, L_1\right)}\right) \end{aligned} \tag{59}$$

where the second inequality follows from the convexity of $\mathcal{P}_{\mathbf{e}}(M, \mathsf{n}, L, P_{\mathbf{x}})$ in $P_{\mathbf{x}}$.
Now consider a code which uses the first n_1 time units of the original encoding scheme as its encoding scheme. Decoder of this new code draws a real number from $[0,1]$ uniformly at random, independently of $\mathsf{Y}^{\mathsf{n}_1}$ (and the message evidently). If this number is less than s it declares erasure else it makes a maximum likelihood decoding with list of size L_1. Then the sum on the left hand side of the below expression (60) is its error probability. But that probability is lower bounded by $\mathcal{P}_{\mathbf{e}}\left(M, \mathsf{n}_1, L_1, s\right)$ which is minimum error probability over all length n_1 block-codes with M messages and list size L_1, i.e.

$$\sum\nolimits_{y^{\mathsf{n}_1}} \mathbf{P}[y^{\mathsf{n}_1}]\,(1-s)\Omega\left(\varphi_{y^{\mathsf{n}_1}}, L_1\right) \geq \mathcal{P}_{\mathbf{e}}\left(M, \mathsf{n}_1, L_1, s\right). \tag{60}$$

Then the theorem follows from the equations (59) and (60) and the fact that $P_{\mathbf{e}}(M, \mathsf{n}, L_1, P_{\mathbf{x}})$ is decreasing function of $P_{\mathbf{x}}$.

QED

Like the result of Shannon, Gallager and Berlekamp in [29, Theorem 1], Theorem 2 is correct both with and without feedback. Although $\mathcal{P}_{\mathbf{e}}$'s are different in each case, the relationship between them given in equation (46) holds in both cases.

B. Classic Results on Error Exponent of Erasure-free block-code with Feedback:

In this section we give a very brief overview of the previously known result on the error probability of erasure free block-codes with feedback. These result are used in Section IV-C together with Theorem 2 to bound $\mathcal{E}_{\mathbf{e}}(R, E_{\mathbf{x}})$ from above. Note that Theorem 2 only relates the error probability of longer codes to that of the shorter ones. It does not in and of itself bound the error probability. It is in a sense a tool to glue together various bounds on the error probability.

First bound we consider is on the error exponent of erasure free block-codes with feedback. Haroutunian proved in [16] that, for any $(M_{\mathsf{n}}, \mathsf{n}, L_{\mathsf{n}})$ sequence of triples, such that $\lim_{\mathsf{n}\to\infty} \frac{\ln M_{\mathsf{n}} - \ln L_{\mathsf{n}}}{\mathsf{n}} = R$,

$$\lim_{\mathsf{n}\to\infty} \tfrac{-\ln \mathcal{P}_{\mathbf{e}}(M_{\mathsf{n}}, \mathsf{n}, L_{\mathsf{n}}, 0)}{\mathsf{n}} \leq E_H(R) \tag{61}$$

where

$$E_H(R) = \min_{V:\mathcal{C}(V)\leq R} \max_{P} \mathsf{D}\left(V \| W|P\right) \qquad \text{and} \qquad \mathcal{C}(V) = \max_{P} \mathsf{I}\left(P, V\right). \tag{62}$$

Second bound we consider is on the tradeoff between the error exponents of two messages in a two message erasure free block-code with feedback. Berlekamp mentions this result while passing in [1] and attributes it to Gallager and Shannon.

Lemma 3: For any feedback encoding scheme with two messages and erasure free decision rule and $T \geq T_0$:

$$\text{either} \qquad P_{\mathbf{e}1} \geq \tfrac{1}{4} e^{-\mathsf{n}T + \sqrt{\mathsf{n}} 4 \ln P_{min}} \qquad \text{or} \qquad P_{\mathbf{e}2} > \tfrac{1}{4} e^{-\mathsf{n}\Gamma(T) + \sqrt{\mathsf{n}} 4 \ln P_{min}}. \tag{63}$$

where $P_{min} = \min_{x,y:W(y|x)} W(y|x)$

$$T_0 \triangleq \max_{x,\tilde{x}} - \ln \sum\nolimits_{y:W(y|\tilde{x})>0} W(y|x) \tag{64}$$

$$\Gamma(T) \triangleq \max_{\Pi} \Gamma(T, \Pi). \tag{65}$$

Result is old and somewhat intuitive to those who are familiar with the calculations in the non-feedback case. Thus probably it has been proven a number of times. But we are not aware of a published proof, hence we have included one in Appendix A.

Although Lemma 3 establishes only the converse part $(T, \Gamma(T))$ is indeed the optimal tradeoff for the error exponent of two messages in an erasure free block-code, both with and without feedback. Achievability of this tradeoff has already been established in [29, Theorem 5] for the case without feedback; evidently this implies the achievability with feedback. Furthermore T_0 does have an operational meaning, it is the maximum error exponent first message can have, while the second message has zero error probability. This fact is also proved in Appendix A.

For some channels Lemma 3 gives us a bound on the error exponent of erasure free-codes at zero rate, which is tighter than Haroutunian's bound at zero rate. In order to see this let us first define T^* to be

$$T^* = \max_T \min\{T, \Gamma(T)\}.$$

Note that T^* is finite iff $\sum_y W(y|x)W(y|\tilde{x}) > 0$ for all x, $\tilde{x}$ pairs. Recall that this is also the necessary and sufficient condition of zero-error capacity, $\mathcal{C}_0$, to be zero. $E_H(R)$ on the other hand is infinite for all $R \leq R_\infty$ like $E_{sp}(R)$ where R_∞ is given by,

$$R_\infty = -\min_{P(\cdot)} \max_y \ln \sum\nolimits_{x:W(y|x)>0} P(x)$$

Even in the cases where $E_H(0)$ is finite, $E_H(0) \geq T^*$. We can use this fact, Lemma 3, and Theorem 2, or [29, Theorem 1] for that matter, to strengthen Haroutunian bound at low rates, as follows.

Lemma 4: For all channels with zero zero-error capacity, $\mathcal{C}_0 = 0$ and any sequence of M_n, such that $\lim_{\mathsf{n}\to\infty} \frac{\ln M_\mathsf{n}}{\mathsf{n}} = R$,

$$\lim_{\mathsf{n}\to\infty} \frac{-\ln \mathcal{P}_{\mathbf{e}}(M_\mathsf{n}, \mathsf{n}, 1, 0)}{\mathsf{n}} \leq \tilde{E}_H(R) \tag{66}$$

where

$$\tilde{E}_H(R) = \left\{ \begin{array}{lll} E_H(R) & \text{if} & R \geq R_{ht} \\ T^* + \frac{E_H(R_{ht}) - T^*}{R_{ht}} R & \text{if} & R \in [0, R_{ht}) \end{array} \right\}$$

and R_{ht} is the unique solution of the equation $T^* = E_H(R) - RE'_H(R)$ if it exists $\mathcal{C}$ otherwise.
Before going into the proof let us note that $\tilde{E}_H(R)$ is obtained simply by drawing the tangent line to the curve $(R, E_H(R))$ from the point $(0, T^*)$. The curve $(R, \tilde{E}_H(R))$ is same as the tangent line, for the rates between 0 and R_{ht}, and it is same as the curve $(R, E_H(R))$ from then on where R_{ht} is the rate of the point at which the tangent from $(0, T^*)$ meets the curve $(R, E_H(R))$.

Proof: For $R \geq R_{ht}$ this Lemma immediately follows from Haroutunian's result in [16] for $L_1 = 1$. If $R < R_{ht}$ then we apply Theorem 2.

$$\mathcal{P}_{\mathbf{e}}(M, \mathsf{n}, L_1, P_{\mathbf{x}}) \geq \mathcal{P}_{\mathbf{e}}(M, \tilde{\mathsf{n}}, L_1, s) \mathcal{P}_{\mathbf{e}} \left(L_1 + 1, \mathsf{n} - \tilde{\mathsf{n}}, \tilde{L}, \tfrac{(1-s)P_{\mathbf{x}}}{\mathcal{P}_{\mathbf{e}}(M, \mathsf{n}, L_1, s)} \right) \tag{67}$$

with[14] $s = 0$, $P_{\mathbf{x}} = 0$, $L_1 = 1$ and $\tilde{\mathsf{n}} = \lfloor \frac{R}{R_{ht}} \rfloor$. on the other hand as a result of Lemma 3 and definition of T^* we have,

$$\mathcal{P}_{\mathbf{e}}(2, \mathsf{n} - \tilde{\mathsf{n}}, L, 0) \geq \frac{e^{-(\mathsf{n}-\tilde{\mathsf{n}})T^* + \sqrt{\mathsf{n}-\tilde{\mathsf{n}}} \ln P_{min}}}{8} \tag{68}$$

[14]Or [29, Theorem 1] with $L_1 = 1$ and $\mathsf{n}_1 = \lfloor \frac{R}{R_{ht}} \rfloor$.

Using equations (67) and (68) we get,

$$\tfrac{-\ln \mathcal{P}_{\mathbf{e}}(M,\mathsf{n},1,0)}{\mathsf{n}} \leq \tfrac{-\ln \mathcal{P}_{\mathbf{e}}(M,\tilde{\mathsf{n}},1,0)}{\tilde{\mathsf{n}}} \tfrac{R}{R_{ht}} + \left[1-\tfrac{R}{R_{ht}}+\tfrac{1}{\tilde{\mathsf{n}}}\right] T^{*} + \left(\sqrt{\tfrac{1}{\tilde{\mathsf{n}}}}\right)\left(\sqrt{\tfrac{R_{ht}-R}{R_{ht}}}\right) \ln \tfrac{P_{min}}{8}$$

where $\frac{\ln M_{\mathsf{n}}}{\tilde{\mathsf{n}}} = R_{ht}$. Lemma follows by simply applying Haroutunian's result to the first terms on the right hand side. ■

C. Generalized Straight Line Bound for Error-Erasure Exponents

Theorem 2 bounds the minimum error probability length n block-codes from below in terms of the minimum error probability of length n_1 and length $(\mathsf{n}-\mathsf{n}_1)$ block-codes. The rate and erasure probability of the longer code constraints the rates and erasure probabilities of the shorter ones, but does not specify them completely. We use this fact together with the improved Haroutunian's bound on the error exponents of erasure free block-codes with feedback, i.e. Lemma 4, and the error exponent tradeoff of the erasure free feedback block-codes with two messages, i.e. Lemma 3, to obtain a family of upper bounds on the error exponents of feedback block-codes with erasure.

Theorem 3: For any DMC with $\mathcal{C}_0 = 0$ rate $R \in [0, \mathcal{C}]$ and $E_{\mathbf{x}} \in [0, E_H(R)]$ and for any $r \in [r_h(R, E_{\mathbf{x}}), \mathcal{C}]$

$$\mathcal{E}_{\mathbf{e}}(R, E_{\mathbf{x}}) \leq \tfrac{R}{r}\tilde{E}_H(r) + (1-\tfrac{R}{r})\Gamma\left(\tfrac{E_{\mathbf{x}}-\frac{R}{r}\tilde{E}_H(r)}{1-\frac{R}{r}}\right)$$

where $r_h(R, E_{\mathbf{x}})$, is the unique solution of $R\tilde{E}_H(r) - rE_{\mathbf{x}} = 0$.

Theorem 3 simply states that any line connecting any two points of the curves $(R, E_{\mathbf{x}}, E_{\mathbf{e}}) = (R, \tilde{E}_H(R), \tilde{E}_H(R))$ and $(R, E_{\mathbf{x}}, E_{\mathbf{e}}) = (0, E_{\mathbf{x}}, \Gamma(E_{\mathbf{x}}))$ lays above the surface $(R, E_{\mathbf{x}}, E_{\mathbf{e}}) = (R, E_{\mathbf{x}}, \mathcal{E}_{\mathbf{e}}(R, E_{\mathbf{x}}))$. The condition $\mathcal{C}_0 = 0$ is not merely a technical condition due to the proof technique; as we will see in Section V for channels with $\mathcal{C}_0 >$, there are zero-error codes with erasure exponent as high as $E_{sp}(R)$ for any rate $R \leq \mathcal{C}$.

Proof: Let us consider Theorem 2, for $s=0$, $L=1$, $L_1=1$, take the logarithm of both sides of equation (46) and divide by n,

$$\tfrac{-\ln \mathcal{P}_{\mathbf{e}}(M,\mathsf{n},1,P_{\mathbf{x}})}{\mathsf{n}} \leq \left(\tfrac{\mathsf{n}_1}{\mathsf{n}}\right) \tfrac{-\ln \mathcal{P}_{\mathbf{e}}(M,\mathsf{n}_1,1,0)}{\mathsf{n}_1} + \left(1-\tfrac{\mathsf{n}_1}{\mathsf{n}}\right) \tfrac{-\ln \mathcal{P}_{\mathbf{e}}\left(2,\mathsf{n}-\mathsf{n}_1,1,\frac{P_{\mathbf{x}}}{\mathcal{P}_{\mathbf{e}}(M,\mathsf{n}_1,1,0)}\right)}{\mathsf{n}-\mathsf{n}_1} \tag{69}$$

Let us assume for the moment that,

$$\mathcal{P}_{\mathbf{e}}(2,\mathsf{n},1,P_{\mathbf{x}}) \geq \tfrac{e^{-\sqrt{\mathsf{n}}\ln P_{min}}}{16} e^{-\mathsf{n}\Gamma\left(\frac{-\ln P_{\mathbf{x}}}{\mathsf{n}} - \frac{\ln 16}{\mathsf{n}} + \frac{\ln P_{min}}{\sqrt{\mathsf{n}}}\right)} \tag{70}$$

where P_{min} is the minimum non-zero entry of W.

If we set $M = \lfloor e^{\mathsf{n}R} \rfloor$, $P_{\mathbf{x}} = e^{-\mathsf{n}E_{\mathbf{x}}}$, $\mathsf{n}_1 = \lfloor \frac{R}{r}\mathsf{n} \rfloor$ in (69), use the bound given in (70) and take the limit as n goes to infinity we get

$$\mathcal{E}_{\mathbf{e}}(R, E_{\mathbf{x}}) \leq \tfrac{R}{r}\mathcal{E}_{\mathbf{e}}(r) + (1-\tfrac{R}{r})\Gamma\left(\tfrac{rE_{\mathbf{x}}-R\mathcal{E}_{\mathbf{e}}(r)}{r-R}\right).$$

Then Theorem 3 follows from Lemma 4 and the fact that $\Gamma(T)$ is nondecreasing function of T.

In order to finish the proof we need to prove (70). Note that if

$$\tfrac{-\ln P_{\mathbf{x}}}{\mathsf{n}} - \tfrac{\ln 16}{\mathsf{n}} + \tfrac{\ln P_{min}}{\sqrt{\mathsf{n}}} \leq T_0$$

then the claim holds trivially, because $\Gamma(T) = \infty$ for $T \leq T_0$. For the case

$$\tfrac{-\ln P_{\mathbf{x}}}{\mathsf{n}} - \tfrac{\ln 16}{\mathsf{n}} + \tfrac{\ln P_{min}}{\sqrt{\mathsf{n}}} > T_0$$

We prove the claim by contradiction. Let us assume that what we have claimed is wrong. Then there exists a $T_0 < T \leq T^*$ such that

$$\mathcal{P}_{\mathbf{e}}(M,\mathsf{n},L,P_{\mathbf{x}}) \leq \tfrac{e^{-\mathsf{n}\Gamma(T)+\sqrt{\mathsf{n}}\ln P_{min}}}{16} \qquad \text{and} \qquad P_{\mathbf{x}} = \tfrac{e^{-\mathsf{n}T+\sqrt{\mathsf{n}}\ln P_{min}}}{16}.$$

Then there exists a block-code with erasures that satisfies

$$\mathbf{P}\left[\tilde{\Upsilon}(\tilde{m})\middle|\, \tilde{m}\right] \leq 2\tfrac{e^{-\mathsf{n}\Gamma(T)+\sqrt{\mathsf{n}}\ln P_{min}}}{16} \qquad \mathbf{P}[\Upsilon(\mathbf{x})|\, \tilde{m}] \leq 2\tfrac{e^{-\mathsf{n}T+\sqrt{\mathsf{n}}\ln P_{min}}}{16}$$

$$\mathbf{P}\left[\tilde{\Upsilon}(\tilde{m})\middle|\, \tilde{\tilde{m}}\right] \leq 2\tfrac{e^{-\mathsf{n}\Gamma(T)+\sqrt{\mathsf{n}}\ln P_{min}}}{16} \qquad \mathbf{P}[\Upsilon(\mathbf{x})|\, \tilde{\tilde{m}}] \leq 2\tfrac{e^{-\mathsf{n}T+\sqrt{\mathsf{n}}\ln P_{min}}}{16}.$$

Let us enlarge the decoding region of $\tilde{m}$ by taking its union with the erasure region:

$$\Upsilon'(\tilde{m}) = \Upsilon(\tilde{m}) \cup \Upsilon(\mathbf{x}) \qquad \Upsilon'(\tilde{\tilde{m}}) = \Upsilon(\tilde{\tilde{m}}) \qquad \Upsilon'(\mathbf{x}) = \emptyset$$

The resulting code is an erasure free code with

$$\mathbf{P}\left[\Upsilon'(\tilde{\tilde{m}})\middle|\, \tilde{m}\right] \leq 2\frac{e^{-\mathrm{n}\Gamma(T)+\sqrt{\mathrm{n}}\ln P_{min}}}{16} \qquad \text{and} \qquad \mathbf{P}\left[\Upsilon'(\tilde{m})\middle|\, \tilde{\tilde{m}}\right] \leq 2\frac{e^{-\mathrm{n}\Gamma(T)+\sqrt{\mathrm{n}}\ln P_{min}}}{16} + 2\frac{e^{-\mathrm{n}T+\sqrt{\mathrm{n}}\ln P_{min}}}{16}.$$

Since $T_0 < T \leq T^*$, $\Gamma(T) \geq T$, this contradicts with Lemma 3 thus equation (70) holds. ■

Note that we have set $L_1 = 1$ in the proof but we could have set it to any subexponential function of block length which diverges as n diverges. By doing so we would have replaced $\Gamma(T)$ with $\mathcal{E}_{\mathbf{e}}(0, E_{\mathbf{x}})$, while keeping the term including $\tilde{E}_H(R)$ the same. Since the best known upper bound for $\mathcal{E}_{\mathbf{e}}(0, E_{\mathbf{x}})$ is $\Gamma(E_{\mathbf{x}})$ for $E_{\mathbf{x}} \leq T^*$ final result is same for case with feedback.[15] On the other hand for the case without feedback, which is not the main focus of this paper, this does make a difference. By choosing L_1 to be a subexponential function of block length one can use Telatar's converse result [30, Theorem 4.4] on the error exponent at zero rate and zero erasure exponent without feedback.

In Figure 1, the upper and lower bounds we have derived for error exponent are plotted as a function of erasure exponent for a binary symmetric channel with cross over probability $\epsilon = 0.25$ at rate $R = 8.62 \times 10^{-2}$ nats per channel use. Solid lines are lower bounds to the error exponent for block-codes with feedback, which have established in Section III, and without feedback, which was previously established previously, [13], [9], [30]. Dashed lines are the upper bounds obtained using Theorem 3. Note that all four curves meet at a point on bottom right, this is the point that corresponds to the error exponent of block-codes at rate $R = 8.62 \times 10^{-2}$ nats per channel use and its values are the same with and without feedback since we are on a symmetric channel and our rate is over the critical rate. Any point to the lower right of this point is achievable both with and without feedback.

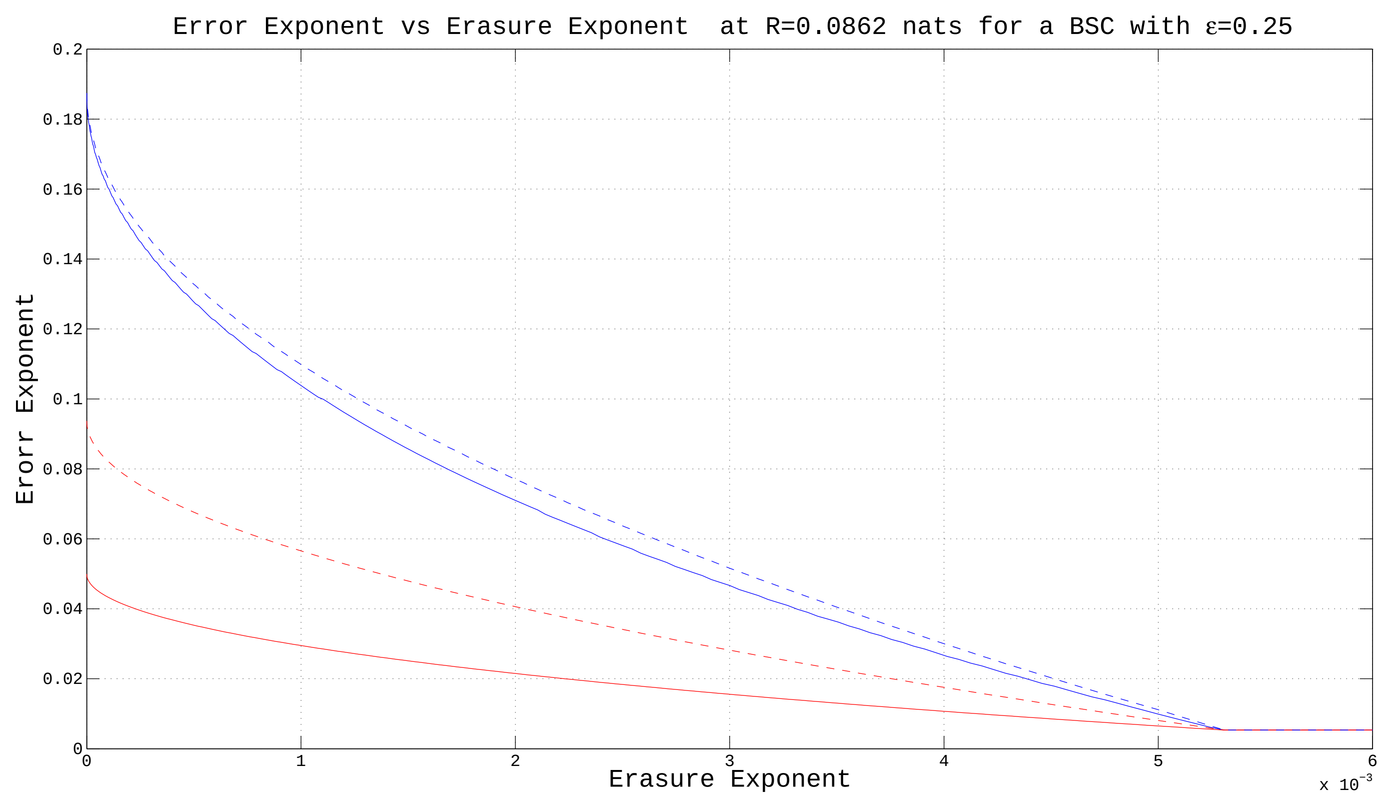

Fig. 1. Error Exponent vs Erasure Exponent

[15]In binary symmetric channels these result can be strengthened using the value of $\mathcal{E}(0)$, [34]. However those changes will improve the upper bound on error exponent only at low rates and high erasure exponents.

V. ERASURE EXPONENT OF ERROR-FREE CODES:$\mathcal{E}_{\mathbf{x}}(R)$

For all DMC's which has one or more zero probability transitions, for all rates below capacity, $R < \mathcal{C}$ and for small enough $E_{\mathbf{x}}$'s, $E_{\mathbf{e}}(R, E_{\mathbf{x}}) = \infty$. For such $(R, E_{\mathbf{x}})$ pairs, coding scheme we have described in Section III gives us an error free code. The connection between the erasure exponent of error free block-codes, and error exponent of block-codes with erasures is not confined to this particular encoding scheme. In order to explain those connections in more detail let us first define the error-free codes more formally.

Definition 3: A sequences $\mathcal{Q}_0$ of block-codes with feedback is an error-free reliable sequence iff

$$P_{\mathbf{e}}{}^{(\mathsf{n})} = 0 \quad \forall \mathsf{n}, \qquad\qquad \text{and} \qquad\qquad \limsup_{\mathsf{n}\to\infty}(P_{\mathbf{x}}{}^{(\mathsf{n})} + \tfrac{1}{|\mathcal{M}^{(\mathsf{n})}|}) = 0.$$

The highest rate achievable for error-free reliable codes is the zero-error capacity with feedback and erasures, $\mathcal{C}_{x,0}$.

If all the transition probabilities are positive i.e. $\min_{x,y} W(y|x) = \delta > 0$, then $\mathbf{P}[y^{\mathsf{n}}|\, m] \geq \delta^n$ for all $m \in \mathcal{M}$ and $y^{\mathsf{n}} \in \mathcal{Y}^{\mathsf{n}}$. Consequently $\mathcal{C}_{x,0}$ is zero. On the other hand as an immediate consequence of the encoding scheme suggested by Yamamoto and Itoh in [32], if there is one or more zero probability transitions, $\mathcal{C}_{x,0}$ is equal to channel capacity $\mathcal{C}$.

Definition 4: For all DMC's with at least one (x, y) pair such that $W(y|x) = 0$, $\forall R < \mathcal{C}$ erasure exponent of error free block-codes with feedback is defined as

$$\mathcal{E}_{\mathbf{x}}(R) \triangleq \sup_{\mathcal{Q}_0 : R(\mathcal{Q}_0) \geq R} E_{\mathbf{x}}(\mathcal{Q}_0). \tag{71}$$

For any erasure exponent, $E_{\mathbf{x}}$ less than $\mathcal{E}_{\mathbf{x}}(R)$, there is an error-free reliable sequence, i.e. there is a reliable sequence with infinite error exponent:

$$E_{\mathbf{x}} \leq \mathcal{E}_{\mathbf{x}}(R) \Rightarrow \mathcal{E}_{\mathbf{e}}(R, E_{\mathbf{x}}) = \infty.$$

More interestingly if $E_{\mathbf{x}} > \mathcal{E}_{\mathbf{x}}(R)$ then $\mathcal{E}_{\mathbf{e}}(R, E_{\mathbf{x}}) < \infty$. In order to see this let δ be the minimum non-zero transition probability. Note that if $\mathbf{P}[y^{\mathsf{n}}|\, m] \neq 0$ then $\mathbf{P}[y^{\mathsf{n}}|\, m] \geq \delta^{\mathsf{n}}$. Thus if $\mathbf{P}\left[\hat{\mathsf{M}} \neq \mathsf{M}\right] \neq 0$, then $\mathbf{P}\left[\hat{\mathsf{M}} \neq \mathsf{M}\right] \geq \delta^{\mathsf{n}} e^{-\mathsf{n}R}$, i.e. $\frac{-\ln P_{\mathbf{e}}{}^{(\mathsf{n})}}{\mathsf{n}} \leq R - \ln \delta$. However if $E_{\mathbf{x}} > \mathcal{E}_{\mathbf{x}}(R)$ then there is no error free reliable sequence at rate R with erasure exponent $E_{\mathbf{x}}$. Thus $P_{\mathbf{e}}{}^{(\mathsf{n})} > 0$ for infinitely many n in any reliable sequence and error exponent of all of those codes are bounded above by a finite number. Consequently,

$$E_{\mathbf{x}} > \mathcal{E}_{\mathbf{x}}(R) \Rightarrow \mathcal{E}_{\mathbf{e}}(R, E_{\mathbf{x}}) < \infty.$$

In a sense like the error exponent of erasure free block-codes, $\mathcal{E}(R)$, erasure exponent of the error free bock codes, $\mathcal{E}_{\mathbf{x}}(R)$, gives a partial description of $\mathcal{E}(R, E_{\mathbf{x}})$. $\mathcal{E}(R)$ gives the value of error exponents below which erasure exponent can be pushed to infinity and $\mathcal{E}_{\mathbf{x}}(R)$ gives the value of erasure exponent below which error exponent can be pushed to infinity.

Below the erasure exponent of zero-error codes, $\mathcal{E}_{\mathbf{x}}(R)$, is investigated separately for two families of channels: Channels which have a positive zero error capacity, i.e. $\mathcal{C}_0 > 0$ and Channels which have zero zero-error capacity, i.e. $\mathcal{C}_0 = 0$.

A. *Case 1: $\mathcal{C}_0 > 0$*

Theorem 4: For a DMC if $\mathcal{C}_0 > 0$ then,

$$E_H(R) \geq \mathcal{E}_{\mathbf{x}}(R) \geq E_{sp}(R).$$

Proof: If zero-error capacity is strictly greater then zero, i.e. $\mathcal{C}_0 > 0$, then one can achieve the sphere packing exponent, with zero error probability using a two phase scheme. In the first phase transmitter uses a length $\mathsf{n}_1 = \lceil e^{\mathsf{n}_1 R} \rceil$ block-code without feedback with a list decoder of size $L = \left\lceil \frac{\partial}{\partial R} E_{sp}(R, P_R^*) \right\rceil$ where P_R^* is the input distribution satisfying $E_{sp}(R) = E_{sp}(R, P_R^*)$. Note that with this list size is sphere packing exponent[16] is achievable at rate R. Thus correct message is in the list with at least probability $(1 - e^{-\mathsf{n}_1 E_{sp}(R)})$, see [9, Page 196]. In the second phase

[16]Indeed this upper bound on error probability is tight exponentially for block-codes without feedback.

transmitter uses a zero error code, of length[17] $\mathsf{n}_2 = \lceil \frac{\ln(L+1)}{\mathcal{C}_0} \rceil$ with $L+1$ messages, to tell the receiver whether the correct message is in that list or not, and the correct message itself if it is in the list. Clearly such a feedback code with two phases is error free, and it has erasures only when there exists an error in the first phase. Thus the erasure probability of the over all code is upper bounded by $e^{-\mathsf{n}_1 E_{sp}(R)}$. Note that n_2 is fixed for a given R. Consequently as the length of the first phase, n_1, diverges, the rate and erasure exponent of $(\mathsf{n}_1 + \mathsf{n}_2)$ long block-code converge to the rate and error exponent of n_1 long code of the first phase, i.e. to R and $E_{sp}(R)$. Thus

$$\mathcal{E}_{\mathbf{x}}(R) \geq E_{sp}(R).$$

Any error free block-code with erasures can be forced to decode, at erasures. The resulting fixed length code has an error probability no larger than the erasure probability of the original code. However we know that, [16], error probability of the erasure free block-codes with feedback decreases with an exponent no larger than $E_H(R)$. Thus,

$$\mathcal{E}_{\mathbf{x}}(R) \leq E_H(R).$$

This upper bound on the erasure exponent also follows from the converse result we present in the next section, Theorem 6. ■

For symmetric channels $E_H(R) = E_{sp}(R)$ and Theorem 4 determines the erasure exponent of error-free codes on symmetric channels with non-zero zero-error-capacity completely.

B. *Case 2:* $\mathcal{C}_0 = 0$

This case is more involved than the previous one. We first establish an upper bound on $\mathcal{E}_{\mathbf{x}}(R)$ in terms of the improved version of Haroutunian's bound, i.e. Lemma 4, and the erasure exponent of error-free codes at zero rate, $\mathcal{E}_{\mathbf{x}}(0)$. Then we show that $\mathcal{E}_{\mathbf{x}}(0)$ is equal to the erasure exponent error-free block-codes with two messages, $\mathcal{E}_{\mathbf{x},2}$, and bound $\mathcal{E}_{\mathbf{x},2}$ from below.

For any M, n and L, $\mathcal{P}_{\mathbf{e}}(M, \mathsf{n}, L, P_{\mathbf{x}}) = 0$ for large enough $P_{\mathbf{x}}$. We denote the minimum of such $P_{\mathbf{x}}$'s by $\mathcal{P}_{0,\mathbf{x}}(M, \mathsf{n}, L)$. Thus we can write $\mathcal{E}_{\mathbf{x},2}$ as

$$\mathcal{E}_{\mathbf{x},2} = \liminf_{\mathsf{n} \to \infty} \mathcal{P}_{0,\mathbf{x}}(2, \mathsf{n}, 1).$$

Theorem 5: For any n, M, L, $\mathsf{n}_1 \leq \mathsf{n}$ and L_1, minimum erasure probability of fixed length error-free block-codes with feedback, $\mathcal{P}_{0,\mathbf{x}}(M, \mathsf{n}, L)$, satisfies

$$\mathcal{P}_{0,\mathbf{x}}(M, \mathsf{n}, L) \geq \mathcal{P}_{\mathbf{e}}(M, \mathsf{n}_1, L_1, 0)\mathcal{P}_{0,\mathbf{x}}(L_1 + 1, \mathsf{n} - \mathsf{n}_1, L). \tag{72}$$

Like Theorem 2, Theorem 5 is correct both with and without feedback. Although $\mathcal{P}_{0,\mathbf{x}}$'s and $\mathcal{P}_{\mathbf{e}}$ will be different in each case, the relationship between them given in equation (72) holds in both cases.

Proof: If $\mathcal{P}_{\mathbf{e}}(M, \mathsf{n}_1, L_1, 0) = 0$ theorem holds trivially. Thus we assume henceforth that $\mathcal{P}_{\mathbf{e}}(M, \mathsf{n}_1, L_1, 0) > 0$. Using Theorem 2 with $P_{\mathbf{x}} = \mathcal{P}_{0,\mathbf{x}}(M, \mathsf{n}, L)$ we get

$$\mathcal{P}_{\mathbf{e}}(M, \mathsf{n}, L, \mathcal{P}_{0,\mathbf{x}}(M, \mathsf{n}, L)) \geq \mathcal{P}_{\mathbf{e}}(M, \mathsf{n}_1, L_1, 0)\mathcal{P}_{\mathbf{e}}\left(L_1 + 1, (\mathsf{n} - \mathsf{n}_1), L, \tfrac{\mathcal{P}_{0,\mathbf{x}}(M,\mathsf{n},L)}{\mathcal{P}_{\mathbf{e}}(M,\mathsf{n}_1,L_1,0)}\right).$$

Since $\mathcal{P}_{\mathbf{e}}(M, \mathsf{n}, L, \mathcal{P}_{0,\mathbf{x}}(M, \mathsf{n}, L)) = 0$ and $\mathcal{P}_{\mathbf{e}}(M, \mathsf{n}_1, L_1, 0) > 0$ we have,

$$\mathcal{P}_{\mathbf{e}}\left(L_1 + 1, (\mathsf{n} - \mathsf{n}_1), L, \tfrac{\mathcal{P}_{0,\mathbf{x}}(M,\mathsf{n},L)}{\mathcal{P}_{\mathbf{e}}(M,\mathsf{n}_1,L_1,0)}\right) = 0.$$

Thus

$$\tfrac{\mathcal{P}_{0,\mathbf{x}}(M,\mathsf{n},L)}{\mathcal{P}_{\mathbf{e}}(M,\mathsf{n}_1,L_1,0)} \geq \mathcal{P}_{0,\mathbf{x}}(L_1 + 1, (\mathsf{n} - \mathsf{n}_1), L).$$

■

As we have done in the errors and erasures case we can convert this into a bound on exponents. If we use the improved version of Haroutunian's bound, i.e. Lemma 4, as an upper bound on the error exponent of erasure free block-codes we get the following.

[17]For some DMCs with $\mathcal{C}_0 > 0$ and for some L one may need more than $\lceil \frac{\ln(L+1)}{\mathcal{C}_0} \rceil$ time units to convey one of the $(L+1)$ messages without any errors, because $\mathcal{C}_0$ itself is defined as a limit. But even in those cases we are guaranteed to have a fixed amount of time for that transmissions, which does not change with n_1. Thus above argument holds as is even in those cases.

Theorem 6: For any rate $R \geq 0$ for any $\alpha \in \left[\frac{R}{\mathcal{C}}, 1\right]$

$$\mathcal{E}_{\mathbf{x}}(R) \leq \alpha \tilde{E}_H\left(\tfrac{R}{\alpha}\right) + (1-\tilde{\alpha})\mathcal{E}_{\mathbf{x}}(0)$$

Now let us focus on the value of erasure exponent at zero rate:

Lemma 5: For the channels which has zero zero-error capacity, i.e. $\mathcal{C}_0 = 0$, erasure exponent of error free block-codes at zero rate $\mathcal{E}_{\mathbf{x}}(0)$ is equal to the erasure exponent of error free block-codes with two messages $\mathcal{E}_{\mathbf{x},2}$.
Note that unlike the two message case, $\mathcal{E}_{\mathbf{x},2}$, in the zero rate case $\mathcal{E}_{\mathbf{x}}(0)$ the number of messages are increasing with block length to infinity, thus we can not claim their equality just as a result of their definitions.

Proof: If we write Theorem 5 for $L = 1$, $\mathsf{n}_1 = 0$ and $L_1 = 1$

$$\begin{aligned}\mathcal{P}_{0,\mathbf{x}}(M, \mathsf{n}, 1) &\geq \mathcal{P}_{\mathbf{e}}(M, 0, 1)\mathcal{P}_{0,\mathbf{x}}(2, \mathsf{n}, 1)\\ &= \tfrac{M-1}{M}\mathcal{P}_{0,\mathbf{x}}(2, \mathsf{n}, 1) \qquad \forall M, \mathsf{n}\end{aligned}$$

Thus as an immediate result of the definitions of $\mathcal{E}_{\mathbf{x}}(0)$ and $\mathcal{E}_{\mathbf{x},2}$, we have $\mathcal{E}_{\mathbf{x}}(0) \leq \mathcal{E}_{\mathbf{x},2}$.

In order to prove the equality one needs to prove $\mathcal{E}_{\mathbf{x}}(0) \geq \mathcal{E}_{\mathbf{x},2}$. For doing that let us assume that it is possible to send one bit with erasure probability ϵ with block-code of length $\ell(\epsilon)$:

$$\epsilon \geq \mathcal{P}_{0,\mathbf{x}}(2, \ell(\epsilon), 1) \tag{73}$$

One can use this code to send r bits, by repeating each bit whenever there exists an erasure. If the block length is $\mathsf{n} = k\ell(\epsilon)$ then a message erasure occurs only when the number of bit erasures in k trials is more then $k - r$. Let $\#e$ denote the number of erasures out of k trials then

$$\mathbf{P}[\#e = l] = \tfrac{k!}{(k-l)!l!}(1-\epsilon)^{k-l}\epsilon^l \qquad \text{and} \qquad P_{\mathbf{x}} = \sum\nolimits_{l=k-r+1}^{k} \mathbf{P}[\#e = l].$$

Thus

$$\begin{aligned}P_{\mathbf{x}} &= \sum\nolimits_{l=k-r+1}^{k} \tfrac{k!}{l!(k-l)!}(1-\epsilon)^{k-l}\epsilon^l\\ &= \sum\nolimits_{l=k-r+1}^{k} \tfrac{k!}{l!(k-l)!}\left(\tfrac{l}{k}\right)^l\left(1-\tfrac{l}{k}\right)^{k-l} e^{-[l\ln\frac{l/k}{\epsilon} + (k-l)\ln\frac{1-l/k}{1-\epsilon}]}\\ &= \sum\nolimits_{l=k-r+1}^{k} \tfrac{k!}{l!(k-l)!}\left(\tfrac{l}{k}\right)^l\left(1-\tfrac{l}{k}\right)^{k-l} e^{-k\mathsf{D}\left(\frac{l}{k}\|\epsilon\right)}.\end{aligned}$$

Then for any $\epsilon \leq 1 - \frac{r}{k}$, we have

$$P_{\mathbf{x}} \leq e^{-k\mathsf{D}\left(1-\frac{r}{k}\|\epsilon\right)}.$$

Evidently $P_{\mathbf{x}} \geq \mathcal{P}_{0,\mathbf{x}}(2^r, \mathsf{n}, 1)$ for $\mathsf{n} = k\ell(\epsilon)$. Thus,

$$\frac{-\ln \mathcal{P}_{0,\mathbf{x}}(2^r, \mathsf{n}, 1)}{\mathsf{n}} \geq \frac{\mathsf{D}\left(1-\frac{r}{k}\|\epsilon\right)}{\ell(\epsilon)}.$$

Then for any sequence of (r, k)'s such that $\lim_{k\to\infty}\frac{r}{k} = 0$ we have $\mathcal{E}_{\mathbf{x}}(0) \geq \frac{-\ln\epsilon}{\ell(\epsilon)}$. Thus any exponent achievable for two message case is achievable for zero rate case: $\mathcal{E}_{\mathbf{x}}(0) \geq \mathcal{E}_{\mathbf{x},2}$. ■

As a result of Lemma 6 which is presented in the next section we know that

$$\mathcal{P}_{0,\mathbf{x}}(2, \mathsf{n}, 1) \geq \tfrac{1}{2}(\sup_{s\in(0,.5)} \beta(s))^{\mathsf{n}} \quad \text{where} \quad \beta(s) = \min_{x,\tilde{x}} \sum\nolimits_y W(y|x)^{(1-s)}W(y|\tilde{x})^s.$$

Thus as a result of Lemma 5 we have

$$\mathcal{E}_{\mathbf{x}}(0) = \mathcal{E}_{\mathbf{x},2} \leq -\ln \sup_{s\in(0,.5)} \beta(s).$$

C. *Lower Bounds on* $\mathcal{P}_{0,\mathbf{x}}(2,\mathsf{n},1)$

Suppose at time t the correct message, M, is assigned to the input letter x and the other message is assigned to the input letter $\tilde{x}$, then the receiver can not to rule out the incorrect message at time t with probability $(\sum_{y:W(y|\tilde{x})>0} WW(y|x)$. Using this fact one can prove that,

$$\mathcal{P}_{0,\mathbf{x}}(2,\mathsf{n},1) \geq \left(\min_{x,\tilde{x}} \sum\nolimits_{y:W(y|\tilde{x})} W(y|x)\right)^{\mathsf{n}}. \tag{74}$$

Now let us consider channels whose transition probability matrix W is of the form

$$W = \begin{bmatrix} 1-q & q \\ 0 & 1 \end{bmatrix}$$

Let us denote the output symbol reachable from both of the input letters by $\tilde{y}$. If Y^{n} is a sequence of $\tilde{y}$'s then the receiver can not decode without errors, i.e. it has to declare and erasure. Thus

$$\begin{aligned}\mathcal{P}_{0,\mathbf{x}}(2,\mathsf{n},1) &\geq \tfrac{1}{2}(\mathbf{P}[\mathsf{Y}^{\mathsf{n}} = \tilde{y}\tilde{y}\ldots\tilde{y}|\,\mathsf{M}=1] + \mathbf{P}[\mathsf{Y}^{\mathsf{n}} = \tilde{y}\tilde{y}\ldots\tilde{y}|\,\mathsf{M}=2]) \\ &\overset{(a)}{\geq} \sqrt{\mathbf{P}[\mathsf{Y}^{\mathsf{n}} = \tilde{y}\tilde{y}\ldots\tilde{y}|\,\mathsf{M}=1]\,\mathbf{P}[\mathsf{Y}^{\mathsf{n}} = \tilde{y}\tilde{y}\ldots\tilde{y}|\,\mathsf{M}=2]} \\ &\overset{(b)}{\geq} q^{\frac{\mathsf{n}}{2}}\end{aligned}$$

where (a) hods because arithmetic mean is larger than the geometric mean, and (b) holds because

$$\mathbf{P}\left[\mathsf{Y}_t = \tilde{y}|\,\mathsf{M}=1,\mathsf{Y}^{t-1}\right]\mathbf{P}\left[\mathsf{Y}_t = \tilde{y}|\,\mathsf{M}=2,\mathsf{Y}^{t-1}\right] \geq q \qquad \forall t$$

Indeed this is bound is tight. If the encoder assigns first message to the input letter that always leads to $\tilde{y}$ and the second message to the other input letter in first $\lfloor\frac{\mathsf{n}}{2}\rfloor$ time instances, and does the flipped assignment in the last $\lceil\frac{\mathsf{n}}{2}\rceil$ time instances, then an erasure happens with a probability less than $q^{\lfloor\frac{\mathsf{n}}{2}\rfloor}$.

Note that equation (74) bounds $\mathcal{P}_{0,\mathbf{x}}(2,\mathsf{n},1)$ only by q^{n}, rather than $q^{\lfloor\frac{\mathsf{n}}{2}\rfloor}$. Using the insight from this example one can establish the following lower bound,

$$\mathcal{P}_{0,\mathbf{x}}(2,\mathsf{n},1) \geq \tfrac{1}{2}\left(\min_{x,\tilde{x}} \sum\nolimits_{y} \sqrt{W(y|x)W(y|\tilde{x})}\right)^{\mathsf{n}}. \tag{75}$$

However the bound given in equation (75) is decaying exponentially in n, even when all entries of the W are positive, i.e. even when $\mathcal{P}_{0,\mathbf{x}}(2,\mathsf{n},1) = 1$. In other words it is not superior to the bound given in equation (74). Following bound implies bounds given in equations (74) and (75). Furthermore for certain channels it is strictly better than both.

Lemma 6: Erasure probability of a error free codes two messages is lower bounded as

$$\mathcal{P}_{0,\mathbf{x}}(2,\mathsf{n},1) \geq \tfrac{1}{2}(\sup_{s\in(0,.5)} \beta(s))^{\mathsf{n}} \quad \text{where} \quad \beta(s) = \min_{x,\tilde{x}} \sum\nolimits_{y} W(y|x)^{(1-s)}W(y|\tilde{x})^{s} \tag{76}$$

Note that the bound in equation (74) is implied by $\lim_{s\to 0^+}\beta(s)$ case, and bound in equation (75) is implied by $\lim_{s\to 0.5^-}\beta(s)$.

Although $\sum_y W(y|x)^s W(y|\tilde{x})^{1-s}$ is convex in s on $(0,0.5)$ for all $(x,\tilde{x})$ pairs, $\beta(s)$ is not convex in s because of the minimization in its definition. Thus the supremum over s does not necessarily occur on the boundaries. Indeed there are channels for which bound given in Lemma 6 is strictly better than the bounds given in (74) and (75). Following is the transition probabilities of such one such channel.

$$W = \begin{bmatrix} 0.1600 & 0.0200 & 0.2200 & 0.3000 & 0.3000 \\ 0.0900 & 0.4000 & 0.2700 & 0.0002 & 0.2398 \\ 0.1800 & 0.2000 & 0.3000 & 0.3200 & 0 \end{bmatrix}$$

$\beta(0) = 0.7,\ \beta(.5) = 0.7027,\ \beta(0.18) = 0.7299.$

Proof: For any error free code and for any $s \in (0, 0.5)$

$$
\begin{aligned}
P_{\mathbf{x}} &= \tfrac{1}{2} \sum_{y^{\mathsf{n}}:\mathbf{P}[y^{\mathsf{n}}|\mathsf{M}=1]\mathbf{P}[y^{\mathsf{n}}|\mathsf{M}=2]>0} \mathbf{P}[y^{\mathsf{n}}|\,\mathsf{M}=1] \left(1+\tfrac{\mathbf{P}[y^{\mathsf{n}}|\mathsf{M}=2]}{\mathbf{P}[y^{\mathsf{n}}|\mathsf{M}=1]}\right) \\
&\geq \tfrac{1}{2} \sum_{y^{\mathsf{n}}:\mathbf{P}[y^{\mathsf{n}}|\mathsf{M}=1]\mathbf{P}[y^{\mathsf{n}}|\mathsf{M}=2]>0} \mathbf{P}[y^{\mathsf{n}}|\,\mathsf{M}=1] \left(1+\tfrac{\mathbf{P}[y^{\mathsf{n}}|\mathsf{M}=2]}{\mathbf{P}[y^{\mathsf{n}}|\mathsf{M}=1]}\right)^{s} \\
&\geq \tfrac{1}{2} \sum_{y^{\mathsf{n}}:\mathbf{P}[y^{\mathsf{n}}|\mathsf{M}=1]\mathbf{P}[y^{\mathsf{n}}|\mathsf{M}=2]>0} \mathbf{P}[y^{\mathsf{n}}|\,\mathsf{M}=1]^{1-s}\,\mathbf{P}[y^{\mathsf{n}}|\,\mathsf{M}=2]^{s} \\
&= \tfrac{1}{2} \sum_{y^{\mathsf{n}-1}:\mathbf{P}[y^{\mathsf{n}-1}|\mathsf{M}=1]\mathbf{P}[y^{\mathsf{n}-1}|\mathsf{M}=2]>0} \mathbf{P}\left[y^{\mathsf{n}-1}\middle|\,\mathsf{M}=1\right]^{1-s} \mathbf{P}\left[y^{\mathsf{n}-1}\middle|\,\mathsf{M}=2\right] \sum_{y_{\mathsf{n}}:\mathbf{P}[y^{\mathsf{n}}|\mathsf{M}=1]\mathbf{P}[y^{\mathsf{n}}|\mathsf{M}=2]>0} \mathbf{P}\left[y_{\mathsf{n}}\middle|\,\mathsf{M}=1,y^{\mathsf{n}-1}\right]^{1-s} \mathbf{P}\left[y_{\mathsf{n}}\middle|\,\mathsf{M}=2,y^{\mathsf{n}-1}\right] \\
&= \tfrac{1}{2} \sum_{y^{\mathsf{n}-1}:\mathbf{P}[y^{\mathsf{n}-1}|\mathsf{M}=1]\mathbf{P}[y^{\mathsf{n}-1}|\mathsf{M}=2]>0} \mathbf{P}\left[y^{\mathsf{n}-1}\middle|\,\mathsf{M}=1\right]^{1-s} \mathbf{P}\left[y^{\mathsf{n}-1}\middle|\,\mathsf{M}=2\right] \beta(s) \\
&= \frac{1}{2}(\beta(s))^{\mathsf{n}}
\end{aligned}
$$

Lemma follows by taking the supremum over $s \in (0, 0.5)$. ∎

VI. Discussion

In the erasure-free case, the error exponent is not known for a general DMC. We do not even know if it is still upper bounded by sphere packing exponent for non-symmetric DMCs. However for the case with erasures, at zero erasure exponent, the value of error exponent been known for long, [3], [32]. Our main aim was establishing upper and lower bounds that will extend the bounds at the zero erasure exponent case gracefully and non-trivially to the positive exponents. Our results are best understood in this framework and should be interpreted accordingly.

We derived inner bounds using a two phase encoding schemes, which are known to be optimal at zero-erasure exponent case. We have improved their performance at positive erasure exponent values by choosing relative durations of the phases properly and by using an appropriate decoder. However within each phase the assignment of messages to input letters is fixed. In a general feedback encoder, on the other hand, assignment of the messages to input symbols at each time can depend on the previous channel outputs and such encoding schemes have proven to improve the error exponent at low rates, [33], [12], [6], [23] for some DMCs. Using such an encoding in the communication phase will improve the performance at low rates. In addition instead of committing to a fixed duration for the communication phase one might consider using a stopping time to switch from communication phase to the control phase. However in order to apply those ideas effectively for a general DMC, it seems one first needs to solve the problem for the erasure-free block-codes for a general DMC.

We derived the outer bounds without making any assumption about the feedback encoding scheme. Thus they are valid for any fixed length block-code with feedback and erasures. The principal idea of the straight line bound is making use of the bounds derived for different rate and erasure exponent points by taking their convex combinations. This approach can be interpreted as a generalization of the outer bounds used for variable length block-codes, [3], [2]. As it was the case for the inner bounds, it seems in order to improve the outer bounds one needs establish outer bounds on some related problem, i.e. on the error exponents of erasure free block-codes with feedback and on the error exponent erasure exponent trade of at zero rate.

Acknowledgement

Authors are thankful to Emre Telatar for his encouragement on the problem, numerous discussions on error-free codes. In particular the observations presented about z-channels are his and Lemma 5 was proved in 2006 summer at Ecole Polytechnique Federale de Lausanne (EPFL). Authors are thankful to Tsachy Weisman, Amos Lapidoth for bringing Shannon- Gallager resulted mentioned in Berlekamp's thesis to their attention, to Anant Sahai for various discussions on communication problems with feedback. Authors are thankful to Robert G. Gallager for various discussion on the encoding scheme presented in Section III, for various discussion on the two message error exponent trade-off and for his suggestions on the manuscript.

APPENDIX

A. The Error Exponent Trade-off for Feedback Encoding Schemes with Two Message and Erasure Free Decoders :

Lemma 7: $\Gamma(T,\Pi)$ defined in equation (35) is also equal to

$$\Gamma(T,\Pi)=\left\{\begin{array}{ccll}\infty & \text{if} & T<\mathsf{D}\left(U_0\|W_a|\Pi\right) & \\ \mathsf{D}\left(U_s\|W_r|\Pi\right) & \text{if} & T=\mathsf{D}\left(U_s\|W_a|\Pi\right) & \text{for some } s\in[0,1]\\ \mathsf{D}\left(U_1\|W_r|\Pi\right) & \text{if} & T>\mathsf{D}\left(U_1\|W_a|\Pi\right) & \end{array}\right\}$$

where

$$U_s(y|x,\tilde{x})=\left\{\begin{array}{ll}\frac{\mathbb{1}_{\{W(y|\tilde{x})>0\}}}{\sum_{\tilde{y}:W(\tilde{y}|\tilde{x})>0}W(\tilde{y}|x)}W(y|x) & \text{if}\quad s=0\\ \frac{W(y|x)^{1-s}W(y|x)^{s}}{\sum_{\tilde{y}}W(\tilde{y}|x)^{1-s}W(\tilde{y}|\tilde{x})^{s}} & \text{if}\quad s\in(0,1)\\ \frac{\mathbb{1}_{\{W(y|x)>0\}}}{\sum_{\tilde{y}:W(\tilde{y}|x)>0}W(\tilde{y}|\tilde{x})}W(y|\tilde{x}) & \text{if}\quad s=1\end{array}\right\}$$

Proof:

$$\begin{aligned}\Gamma(T,\Pi)&=\min_{U:\mathsf{D}(U\|W_a|\Pi)\leq T}\mathsf{D}\left(U\|W_r|\Pi\right)\\ &=\min_{U}\sup_{\lambda>0}\mathsf{D}\left(U\|W_r|\Pi\right)+\lambda(\mathsf{D}\left(U\|W_a|\Pi\right)-T)\\ &\overset{(a)}{=}\sup_{\lambda>0}\min_{U}\mathsf{D}\left(U\|W_r|\Pi\right)+\lambda(\mathsf{D}\left(U\|W_a|\Pi\right)-T)\\ &=\sup_{\lambda>0}\min_{U}-\lambda T+(1+\lambda)\sum\nolimits_{x,\tilde{x},y}\Pi(x,\tilde{x})U(y|x,\tilde{x})\ln\frac{U(y|x,\tilde{x})}{W(y|x)^{\frac{\lambda}{1+\lambda}}W(y|\tilde{x})^{\frac{1}{1+\lambda}}}\\ &\overset{(b)}{=}\sup_{\lambda>0}-\lambda T-(1+\lambda)\sum\nolimits_{x,\tilde{x}}\Pi(x,\tilde{x})\ln\sum_{y}W(y|x)^{\frac{\lambda}{1+\lambda}}W(y|\tilde{x})^{\frac{1}{1+\lambda}}\end{aligned}\tag{77}$$

where (a) follows from convexity of $\mathsf{D}\left(U\|W_r|\Pi\right)+\lambda(\mathsf{D}\left(U\|W_a|\Pi\right)-T)$ in U and linearity (concavity) of it in λ; (b) holds because minimizing U is U_s for $s=\frac{1}{1+\lambda}$. The function on the right hand side of (77) is maximized at a positive and finite λ iff there is a λ such that $\mathsf{D}\left(U_{\frac{1}{1+\lambda}}\middle\|W_a|\Pi\right)=T$. Thus by substituting $\lambda=\frac{1-s}{s}$ we get

$$\Gamma(T,\Pi)=\left\{\begin{array}{ccll}\infty & \text{if} & T<\lim_{s\to0^+}\mathsf{D}\left(U_s\|W_a|\Pi\right) & \\ \lim_{s\to0^+}\mathsf{D}\left(U_s\|W_r|\Pi\right) & \text{if} & T=\lim_{s\to0^+}\mathsf{D}\left(U_s\|W_a|\Pi\right) & \\ \mathsf{D}\left(U_s\|W_r|\Pi\right) & \text{if} & T=\mathsf{D}\left(U_s\|W_a|\Pi\right) & \text{for some } s\in(0,1)\\ \lim_{s\to1^-}\mathsf{D}\left(U_s\|W_r|\Pi\right) & \text{if} & T=\lim_{s\to1^-}\mathsf{D}\left(U_s\|W_a|\Pi\right) & \\ \lim_{s\to1^-}\mathsf{D}\left(U_s\|W_r|\Pi\right) & \text{if} & T>\lim_{s\to1^-}\mathsf{D}\left(U_s\|W_a|\Pi\right) & \end{array}\right\}\tag{78}$$

Lemma follows from the definition U_s at $s=0,1$ and equation (78). ■

Now we are ready to present the proof of Lemma 3

Proof [Lemma 3]:
Our proof is very much like the one for the converse part of [29, Theorem 5], except few modifications that allow us to handle the fact that encoding schemes we are considering are feedback encoding schemes. Like [29, Theorem 5] we construct a probability measure $P_T[\cdot]$ on $\mathcal{Y}^{\mathsf{n}}$ as a function of T and the encoding scheme. Then we bound the error probability of each message from below using the probability of the decoding region of the other message under $P_T[\cdot]$.

For any $T\geq T_0$ and Π, let $S_{T,\Pi}$ be

$$S_{T,\Pi}=\left\{\begin{array}{ccl}0 & \text{if} & T<\mathsf{D}\left(U_0\|W_a|\Pi\right)\\ s & \text{if} & \exists s\in[0,1]\text{s.t. }\mathsf{D}\left(U_s\|W_a|\Pi\right)=T\\ 1 & \text{if} & T>\mathsf{D}\left(U_1\|W_a|\Pi\right)\end{array}\right\}.\tag{79}$$

Recall that

$$T_0=\max_{x,\tilde{x}}-\ln\sum\nolimits_{y:W(y|\tilde{x})>0}W(y|x)\qquad\text{and}\qquad\mathsf{D}\left(U_0\|W_a|\Pi\right)=-\sum\nolimits_{x,\tilde{x}}\Pi(x,\tilde{x})\ln\sum\nolimits_{y:W(y|\tilde{x})>0}W(y|x).$$

Thus

$$T_0 \geq \mathsf{D}\left(U_0 \| W_a | \Pi \right) \tag{80}$$

Thus as a result of definition of $S_{T,\Pi}$ and equation (80) we have

$$\mathsf{D}\left(U_{S_{T,\Pi}} \middle\| W_a | \Pi \right) \leq T \qquad \forall T \geq T_0. \tag{81}$$

Using Lemma 7, definition of $S_{T,\Pi}$ and equation (80) we can also conclude that

$$\mathsf{D}\left(U_{S_{T,\Pi}} \middle\| W_r | \Pi \right) = \Gamma\left(T, \Pi\right) \leq \Gamma\left(T\right) \qquad \forall T \geq T_0. \tag{82}$$

Note that for a feedback encoding schemes with two messages at time t, $X_t(\cdot) : \{m_1, m_2\} \times \mathcal{Y}^{t-1}$, given the the past channel outputs, y^{t-1} channel inputs for each message ($X_t(m_1, y^{t-1})$ and $X_t(m_2, y^{t-1})$) are fixed. Thus there is a corresponding Π:

$$\Pi(x, \tilde{x}) = \left\{ \begin{array}{lll} 0 & \text{if} & (x, \tilde{x}) \neq (X_t(m_1, y^{t-1}), X_t(m_2, y^{t-1})) \\ 1 & \text{if} & (x, \tilde{x}) = (X_t(m_1, y^{t-1}), X_t(m_2, y^{t-1})) \end{array} \right\}. \tag{83}$$

Then for any $T \geq T_0$ let $P_T\left[y_t | \, y^{t-1} \right]$ be

$$P_T\left[y_t | \, y^{t-1} \right] = U_{S_{T,\Pi}}(y_t | X_t(m_1, y^{t-1}), X_t(m_2, y^{t-1})). \tag{84}$$

Note that as a result of equation (81) and equation (82) we have

$$\sum\nolimits_{y_t} P_T\left[y_t | \, y^{t-1} \right] \ln \tfrac{P_T[y_t | y^{t-1}]}{\mathbf{P}[y_t | \mathsf{M} = m_1, y^{t-1}]} \leq T \qquad \text{and} \qquad \sum\nolimits_{y_t} P_T\left[y_t | \, y^{t-1} \right] \ln \tfrac{P_T[y_t | y^{t-1}]}{\mathbf{P}[y_t | \mathsf{M} = m_2, y^{t-1}]} \leq \Gamma\left(T\right)$$

Now we make a standard measure change argument,

$$\begin{aligned} \mathbf{P}[y^{\mathsf{n}} | \, \mathsf{M} = m_1] &= e^{\ln \frac{\mathbf{P}[y^{\mathsf{n}} | \mathsf{M} = m_1]}{P_T[y^{\mathsf{n}}]}} P_T\left[y^{\mathsf{n}} \right] \\ &= e^{-\sum_{t=1}^{\mathsf{n}} \ln \frac{P_T[y_t | y^{t-1}]}{\mathbf{P}[y_t | \mathsf{M} = m_1, y^{t-1}]}} P_T\left[y^{\mathsf{n}} \right] \\ &= e^{-\sum_{t=1}^{\mathsf{n}} \sum_{\tilde{y}_t} P_T[\tilde{y}_t | y^{t-1}] \ln \frac{P_T[\tilde{y}_t | y^{t-1}]}{\mathbf{P}[\tilde{y}_t | \mathsf{M} = m_1, y^{t-1}]}} e^{\sum_{t=1}^{\mathsf{n}} \sum_{\tilde{y}_t} P_T[\tilde{y}_t | y^{t-1}] \left(\ln \frac{P_T[\tilde{y}_t | y^{t-1}]}{\mathbf{P}[\tilde{y}_t | \mathsf{M} = m_1, y^{t-1}]} - \ln \frac{P_T[y_t | y^{t-1}]}{\mathbf{P}[y_t | \mathsf{M} = m_1, y^{t-1}]} \right)} P_T\left[y^{\mathsf{n}} \right] \\ &\leq e^{-\mathsf{n}T} e^{\sum_{t=1}^{\mathsf{n}} Z_{t,1}(y_t | y^{t-1})} P_T\left[y^{\mathsf{n}} \right] \end{aligned} \tag{85}$$

where

$$Z_{t,1}(y_t | y^{t-1}) = \sum_{\tilde{y}_t} P_T\left[\tilde{y}_t | \, y^{t-1} \right] \left(\ln \tfrac{P_T[\tilde{y}_t | y^{t-1}]}{\mathbf{P}[\tilde{y}_t | \mathsf{M} = m_1, y^{t-1}]} - \ln \tfrac{P_T[y_t | y^{t-1}]}{\mathbf{P}[y_t | \mathsf{M} = m_1, y^{t-1}]} \right) \tag{86}$$

Following a very similar reasoning we get,

$$\mathbf{P}[y^{\mathsf{n}} | \, \mathsf{M} = m_2] \geq e^{-\mathsf{n}\Gamma(T)} e^{\sum_{t=1}^{\mathsf{n}} Z_{t,2}(y_t | y^{t-1})} P_T\left[y^{\mathsf{n}} \right] \tag{87}$$

where

$$Z_{t,2}(y_t | y^{t-1}) = \sum_{\tilde{y}_t} P_T\left[\tilde{y}_t | \, y^{t-1} \right] \left(\ln \tfrac{P_T[\tilde{y}_t | y^{t-1}]}{\mathbf{P}[\tilde{y}_t | \mathsf{M} = m_2, y^{t-1}]} - \ln \tfrac{P_T[y_t | y^{t-1}]}{\mathbf{P}[y_t | \mathsf{M} = m_2, y^{t-1}]} \right) \tag{88}$$

Note that for $m = \{m_1, m_2\}$ and $t \in \{1, 2, \ldots, \mathsf{n}\}$,

$$\sum_{y_t} P_T\left[y_t | \, y^{t-1} \right] Z_{t,m}(y_t | y^{t-1}) = 0 \qquad \forall y^{t-1} \in \mathcal{Y}^{t-1} \tag{89a}$$

$$(Z_{t,m}(y_t | y^{t-1}))^2 \leq 4(\ln P_{min})^2 \qquad \forall y^t \in \mathcal{Y}^t \tag{89b}$$

$$\sum_{y_t} P_T\left[y_t | \, y^{t-1} \right] Z_{t,m}(y_t | y^{t-1}) Z_{t-k,m}(y_t | y^{t-1-k}) = 0 \qquad \forall y^{t-1} \in \mathcal{Y}^{t-1} \quad \forall k \{1, 2, \ldots, t-1\} \tag{89c}$$

Thus as a result of equation (89), for all $m = \{m_1, m_2\}$

$$\sum_{y^{\mathsf{n}}} P_T\left[y^{\mathsf{n}}\right] \sum_{t=1}^{\mathsf{n}} Z_{t,m}(y_t|y^{t-1}) = 0 \tag{90a}$$

$$\sum_{y^{\mathsf{n}}} P_T\left[y^{\mathsf{n}}\right] \left(\sum_{t=1}^{\mathsf{n}} Z_{t,m}(y_t|y^{t-1})\right)^2 \leq 4\mathsf{n}(\ln P_{min})^2 \tag{90b}$$

For $m = m_1, m_2$ let $\mathcal{Z}_m$ be

$$\mathcal{Z}_m = \left\{ y^{\mathsf{n}} : |\sum_{t=1}^{\mathsf{n}} Z_{t,m}(y_t|y^{t-1})| \leq 4\sqrt{\mathsf{n}} \ln \tfrac{1}{P_{min}} \right\}$$

Using equation (90) and Chebychev's inequality we conclude that,

$$P_T\left[\mathcal{Z}_m\right] \geq 3/4 \qquad m = m_1, m_2 \Rightarrow P_T\left[\mathcal{Z}_{m_1} \cap \mathcal{Z}_{m_2}\right] \geq 1/2$$

Thus either the total probability of intersection of $\mathcal{Z}_{m_1} \cap \mathcal{Z}_{m_2}$ with the decoding region of second message is equal to or larger than $1/4$ or the total probability of intersection of $\mathcal{Z}_{m_1} \cap \mathcal{Z}_{m_2}$ with the decoding region of first message is strictly larger than $1/4$. Then the lemma follow from equations (85) and (87).

QED

As we have noted previously T_0 does have an operational meaning it is the maximum error exponent first message can have, when the error probability of the second message is zero.

Lemma 8: For any feedback encoding scheme with two messages, if $P_{\mathbf{e}m_2} = 0$ then $P_{\mathbf{e}m_1} \geq e^{-\mathsf{n}T_0}$. Furthermore there does exist an encoding scheme such that $P_{\mathbf{e}m_2} = 0$ then $P_{\mathbf{e}m_1} = e^{-\mathsf{n}T_0}$.

Proof: Let us a similar construction,

$$P_T\left[y_t|\, y^{t-1}\right] = U_0(y_t|X_t(m_1, y^{t-1}), X_t(m_2, y^{t-1})).$$

Recall that

$$U_0(y_t|x, \tilde{x}) = \tfrac{\mathbb{1}_{\{W(y|\tilde{x})>0\}}}{\sum_{\tilde{y}:W(\tilde{y}|\tilde{x})>0} W(\tilde{y}|x)} W(y|x)$$

Thus

$$P_T\left[y_t|\, y^{t-1}\right] \leq e^{T_0} \mathbf{P}\left[y_t|\, \mathsf{M} = m_1, y^{t-1}\right]$$
$$P_T\left[y_t|\, y^{t-1}\right] \leq \mathbb{1}_{\{\mathbf{P}[y_t|\mathsf{M}=m_2, y^{t-1}]>0\}}$$

Then

$$\mathbf{P}[y^{\mathsf{n}}|\, \mathsf{M} = m_1] \geq e^{-\mathsf{n}T_0} P_T\left[y^{\mathsf{n}}\right] \tag{91}$$

$$\mathbf{P}[y^{\mathsf{n}}|\, \mathsf{M} = m_2] \geq e^{\mathsf{n}\ln P_{min}} P_T\left[y^{\mathsf{n}}\right] \tag{92}$$

where P_{min} is the minimum non-zero element of W. Since $P_{\mathbf{e}b} = 0$ equation (92) implies that $P_T\left[\hat{\mathsf{M}} = m_2\right] = 1$. Using this fact and equation (91) we conclude that

$$P_{\mathbf{e}1} \geq e^{-\mathsf{n}T_0}. \tag{93}$$

Let us assume that maximizing x-pair is (x_1^*, x_2^*) i.e. $T_0 = -\ln \sum_{y:W(y|x_2^*)>0} W(y|x_1^*)$. If the the encoding scheme sends x_1^* for the first message and x_2^* for the second message, and the decoder decodes to second message unless $\mathsf{Y}_t = y^*$ for some some $t \in \{1, 2, \ldots, \mathsf{n}\}$ and y^* such that $W(y^*|x_2^*) = 0$ then $P_{\mathbf{e}m_2} = 0$ and $P_{\mathbf{e}m_1} = e^{-\mathsf{n}T_0}$ ∎

B. Convexity of $E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$ in α:

Lemma 9: For any probability distribution P on input alphabet $\mathcal{X}$, $\zeta(P, Q, R)$ is convex in (Q, R) pair.

Proof: Note that

$$\gamma\zeta(R_a, P, Q_a) + (1-\gamma)\zeta(R_b, P, Q_b) = \min_{\substack{V_a, V_b: \begin{smallmatrix} \mathsf{I}(P,V_a) \leq R_a & \mathsf{I}(P,V_b) \leq R_b \\ (PV_a)_Y = Q_a & (PV_b)_Y = Q_b \end{smallmatrix}}} \gamma \mathsf{D}\left(V_a \| W|P\right) + (1-\gamma)\mathsf{D}\left(V_b \| W|P\right)$$

Using the convexity of $\mathsf{D}\left(V\| W|P\right)$ in V and Jensen's inequality we get,

$$\gamma\zeta(R_a, P, Q_a) + (1-\gamma)\zeta(R_b, P, Q_b) \geq \min_{\substack{V_a, V_b: \begin{smallmatrix} \mathsf{I}(P,V_a) \leq R_a & \mathsf{I}(P,V_b) \leq R_b \\ (PV_a)_Y = Q_a & (PV_b)_Y = Q_b \end{smallmatrix}}} \mathsf{D}\left(V_\gamma \| W|P\right)$$

where $V_\gamma = \gamma V_a + (1-\gamma)V_b$.

If the set that a minimization is done over is enlarged, then the resulting minimum does not increase. Using this fact together with the convexity of $\mathsf{I}\left(P, V\right)$ in V and Jensen's inequality we get,

$$\begin{aligned} \gamma\zeta(R_a, P, Q_a) + (1-\gamma)\zeta(R_b, P, Q_b) &\geq \min_{\substack{V_\gamma: \begin{smallmatrix} \mathsf{I}(P,V_\gamma) \leq R_\gamma \\ (PV_\gamma)_Y = Q_\gamma \end{smallmatrix}}} \mathsf{D}\left(V_\gamma \| W|P\right) \\ &= \zeta(R_\gamma, P, Q_\gamma) \end{aligned}$$

where $R_\gamma = \gamma R_a + (1-\gamma)R_b$, $Q_\gamma = \gamma Q_a + (1-\gamma)Q_b$. ■

Lemma 10: For all $(R, E_{\mathbf{x}}, P, \Pi)$ quadruples such that $E_r(R, P) \geq E_{\mathbf{x}}$, $E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$ is a convex function of α on the interval $[\alpha^*(R, E_{\mathbf{x}}, P), 1]$ where $\alpha^*(R, E_{\mathbf{x}}, P)$ is the unique solution[18] of $\alpha E_r(\frac{R}{\alpha}, P) = E_{\mathbf{x}}$.

Proof: For any P such that $E_r(R, P)$ is non-negative, convex and decreasing function of R in the interval $[0, \mathsf{I}\left(P, W\right)]$. Thus $\alpha E_r(\frac{R}{\alpha}, P)$ is strictly increasing continuous function of $\alpha \in [\frac{R}{\mathsf{I}(P,W)}, 1]$. Furthermore for $\alpha = \frac{R}{\mathsf{I}(P,W)}$, $\alpha E_r(\frac{R}{\alpha}, P) = 0$ and for $\alpha = 1$, $\alpha E_r(\frac{R}{\alpha}, P) = 0 \geq E_{\mathbf{x}}$. Thus $\alpha E_r(\frac{R}{\alpha}, P) = E_{\mathbf{x}}$ has a unique solution.

Note that for any $\gamma \in [0, 1]$

$$\begin{aligned} &\gamma E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha_a, P, \Pi) + (1-\gamma)E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha_b, P, \Pi) \\ &\quad = \min_{\substack{Q_a, R_{1a}, R_{2a}, T_a, Q_b, R_{1b}, R_{2b}, T_b: \\ R_{1a} \geq R_{2a} \geq R \quad T_a \geq 0 \\ R_{1b} \geq R_{2b} \geq R \quad T_b \geq 0 \\ \alpha_a \zeta(\frac{R_{1a}}{\alpha_a}, P, Q_a) + R_{2a} - R + T_a \leq E_{\mathbf{x}} \\ \alpha_b \zeta(\frac{R_{1b}}{\alpha_b}, P, Q_b) + R_{2b} - R + T_b \leq E_{\mathbf{x}}}} \begin{array}{l} \gamma\left[\alpha_a \zeta(\frac{R_{2a}}{\alpha_a}, P, Q_a) + R_{1a} - R + (1-\alpha_a)\Gamma\left(\frac{T_a}{1-\alpha_a}, \Pi\right)\right] \\ +(1-\gamma)\left[\alpha_b \zeta(\frac{R_{2b}}{\alpha_b}, P, Q_b) + R_{1b} - R + (1-\alpha_b)\Gamma\left(\frac{T_b}{1-\alpha_b}, \Pi\right)\right] \end{array} \\ &\quad \geq \min_{\substack{Q_\gamma, R_{1\gamma}, R_{2\gamma} T_\gamma: \\ R_{1\gamma} \geq R_{2\gamma} \geq R \quad T_\gamma \geq 0 \\ \alpha_\gamma \zeta(\frac{R_{1\gamma}}{\alpha_\gamma}, P, Q_\gamma) + R_{2\gamma} - R + T_\gamma \leq E_{\mathbf{x}}}} \alpha_\gamma \zeta(\frac{R_{2\gamma}}{\alpha_\gamma}, P, Q_\gamma) + R_{1\gamma} - R + (1-\alpha_\gamma)\Gamma\left(\frac{T}{1-\alpha_\gamma}, \Pi\right) \\ &\quad = E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha_\gamma, P, \Pi). \end{aligned}$$

where α_γ, T_γ, Q_γ, $R_{1\gamma}$ and $R_{2\gamma}$ are given by,

$$\begin{aligned} \alpha_\gamma &= \gamma\alpha_a + (1-\gamma)\alpha_b & T_\gamma &= \gamma T_a + (1-\gamma)T_b & Q_\gamma &= \tfrac{\gamma\alpha_a}{\alpha_\gamma} Q_a + \tfrac{(1-\gamma)\alpha_b}{\alpha_\gamma} Q_b \\ R_{1\gamma} &= \gamma R_{1a} + (1-\gamma)R_{1b} & R_{2\gamma} &= \gamma R_{2a} + (1-\gamma)R_{2b} \end{aligned}$$

The inequality follows from convexity arguments analogous to the ones used in the proof of Lemma 9. ■

[18]The equation $\alpha E_r(\frac{R}{\alpha}, P) = 0$ has multiple solutions; we choose the minimum of those to be the α^* i.e., $\alpha^*(R, 0, P) = \frac{R}{\mathsf{I}(P,W)}$.

C. $\max_{\Pi} E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi) > \max_{\Pi} E_{\mathbf{e}}(R, E_{\mathbf{x}}, 1, P, \Pi), \quad \forall P \in \mathcal{P}\left(R, E_{\mathbf{x}}, \alpha\right)$

Let us first consider a control phase type $\Pi_P(x_1, x_2) = \frac{P(x_1)P(x_2)\mathbb{1}_{\{x_1 \neq x_2\}}}{1-\sum_x (P(x))^2}$ and establish,

$$E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi_P) > E_{\mathbf{e}}(R, E_{\mathbf{x}}, 1, P, \Pi_P) \qquad \forall P \in \mathcal{P}\left(R, E_{\mathbf{x}}, \alpha\right) \tag{94}$$

First consider

$$\begin{aligned}
\mathsf{D}\left(U \| W_a | \Pi_P\right) &= \tfrac{1}{1-\sum_x (P(x))^2} \sum_{x_1, x_2 : x_1 \neq x_2} P(x_1)P(x_2) \sum\nolimits_y U(y|x_1, x_2) \log \tfrac{U(y|x_1,x_2)}{W(y|x_1)} \\
&= \tfrac{1}{1-\sum_x (P(x))^2} \sum_{x_1, x_2 : x_1 \neq x_2} P(x_1)P(x_2) \sum\nolimits_y U(y|x_1, x_2) \left[\log \tfrac{U(y|x_1,x_2)}{V_U(y|x_1)} - \log \tfrac{V_U(y|x_1)}{W(y|x_1)}\right] \\
&\geq \tfrac{1}{1-\sum_x (P(x))^2} \left[\mathsf{I}\left(P, \hat{V}_U\right) + \mathsf{D}\left(V_U \| W | P\right)\right]
\end{aligned} \tag{95}$$

where the last step follows from the log sum inequality and transition probability matrices V_U and $\hat{V}_U$ are given by

$$\begin{aligned}
V_U(y|x_1) &= W(x_1|y)P(x_1) + \sum\nolimits_{x_2 : x_2 \neq x_1} U(y|x_1, x_2)P(x_2) \\
\hat{V}_U(y|x_2) &= W(x_2|y)P(x_2) + \sum\nolimits_{x_1 : x_1 \neq x_2} U(y|x_1, x_2)P(x_1).
\end{aligned}$$

Using a similar line of reasoning we get,

$$\mathsf{D}\left(U \| W_r | \Pi_P\right) \geq \tfrac{1}{1-\sum_x (P(x))^2} \left[\mathsf{D}\left(\hat{V}_U \middle\| W | P\right) + \mathsf{I}\left(P, V_U\right)\right] \tag{96}$$

Note that for all $P \in \mathcal{P}\left(R, E_{\mathbf{x}}, \alpha\right)$ if use the inequalities (95) and (96) together the definition of $E_{\mathbf{e}}$ given in (13) and (18) we get,

$$E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha(R, E_{\mathbf{x}}), P, \Pi_P) \geq E_{\mathbf{e}}(R, E_{\mathbf{x}}, 1, P, \Pi_P) + \delta_P$$

for some $\delta_P > 0$. Consequently for all $P \in \mathcal{P}\left(R, E_{\mathbf{x}}, \alpha\right)$, equation (94) holds.

Note that for all Π and for all $P \in \mathcal{P}\left(R, E_{\mathbf{x}}, \alpha\right)$

$$E_{\mathbf{e}}(R, E_{\mathbf{x}}, 1, P, \Pi_P) = E_{\mathbf{e}}(R, E_{\mathbf{x}}, 1, P, \Pi).$$

Thus we have:

$$\max_{\Pi} E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi) > \max_{\Pi} E_{\mathbf{e}}(R, E_{\mathbf{x}}, 1, P, \Pi) \qquad \forall P \in \mathcal{P}\left(R, E_{\mathbf{x}}, \alpha\right). \tag{97}$$

REFERENCES

[1] E. R. Berlekamp. *Block Coding with Noiseless Feedback.* Ph.d. thesis, Massachusetts Institute of Technology, Department of Electrical Engineering, 1964.

[2] P. Berlin, B. Nakiboğlu, B. Rimoldi, and E. Telatar. A simple converse of Burnashev's reliability function. *Information Theory, IEEE Transactions on*, 55(7):3074–3080, July 2009.

[3] M. V. Burnashev. Data transmission over a discrete channel with feedback, random transmission time. *Problems of Information Transmission*, 12(4):10–30, 1976.

[4] M. V. Burnashev. Note: On the article "data transmission over a discrete channel with feedback, random transmission time". *Problems of Information Transmission*, 13(1):108, 1977.

[5] M. V. Burnashev. Sequential discrimination of hypotheses with control of observations. *Math. USSR Izvestija*, 15(3):419–440, 1980.

[6] M. V. Burnashev. On the reliability function of a binary symmetrical channel with feedback. *Problems of Information Transmission*, 24(1):3–10, 1988.

[7] M. V. Burnashev and H. Yamamoto. Noisy feedback improves the bsc reliability function. pages 1501–1505, 28 2009-July 3 2009.

[8] M. V. Burnashev and H. Yamamoto. On the zero-rate error exponent for a bsc with noisy feedback. *Problems of Information Transmission*, 44(3):198–213, September 2008.

[9] Imre Csiszár and János Körner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press, Inc., Orlando, FL, USA, 1982.

[10] R. L. Dobrushin. An asymptotic bound for the probability error of information transmission through a channel without memory using the feedback. *Problemy Kibernetiki*, 8:161–168, 1962.

[11] S. Draper and A. Sahai. Variable-length coding with noisy feedback. *European Transactions on Telecommunications*, 19(4):355–370, May 2008.

[12] A. G. D'yachkov. Upper bounds on the error probability for discrete memoryless channels with feedback. *Problems of Information Transmission*, 11(4):13–28, 1975.

[13] G. Jr. Forney. Exponential error bounds for erasure, list, and decision feedback schemes. *Information Theory, IEEE Transactions on*, 14(2):206–220, mar 1968.
[14] R. G. Gallager and B. Nakiboğlu. Variations on a theme by Schalkwijk and Kailath. *Information Theory, IEEE Transactions on*, 56(1):6–17, Jan. 2010.
[15] P. K. Gopala, Y-H. Nam, and H. El Gamal. On the error exponents of ARQ channels with deadlines. *Information Theory, IEEE Transactions on*, 53(11):4265–4273, Nov. 2007.
[16] E. A. Haroutunian. A lower bound of the probability of error for channels with feedback. *Problems of Information Transmission*, 13(2):36–44, 1977.
[17] E. Hof, I. Sason, and S. Shamai. Performance bounds for erasure, list and feedback schemes with linear block codes. 2009.
[18] Y-H. Kim, A. Lapidoth, and T. Weissman. Error exponents for the gaussian channel with active noisy feedback. arXiv:0909.4203v1 [cs.IT], http://arxiv.org/abs/0909.4203v1.
[19] N. Merhav. Error exponents of erasure/list decoding revisited via moments of distance enumerators. *Information Theory, IEEE Transactions on*, 54(10):4439–4447, 2008.
[20] N. Merhav and M. Feder. Minimax universal decoding with an erasure option. *Information Theory, IEEE Transactions on*, 53(5):1664–1675, May 2007.
[21] P. Moulin. A Neyman–Pearson approach to universal erasure and list decoding. *Information Theory, IEEE Transactions on*, 55(10):4462–4478, Oct 2009.
[22] B. Nakiboğlu and R. G. Gallager. Error exponents for variable-length block codes with feedback and cost constraints. *IEEE Transactions on Information Theory*, 54(3):945–963, 2008.
[23] B. Nakiboğlu and L Zheng. Upper bounds to error probability with feedback. In *ITA 2010. Information Theory and Applications Workshop*, UCSD CA, 2010.
[24] M. S. Pinsker. The probability of error in block transmission in a memoryless gaussian channel with feedback. *Problems of Information Transmission*, 4(4):1–4, 1968.
[25] E. Sabbag and N. Merhav. Achievable error exponents for channel with side information - erasure and list decoding. arXiv:0903.2203v1 [cs.IT].
[26] J. P. M. Schalkwijk. A coding scheme for additive noise channels with feedback–II: Band-limited signals. *Information Theory, IEEE Transactions on*, 12(2):183–189, 1966.
[27] J. P. M. Schalkwijk and T. Kailath. A coding scheme for additive noise channels with feedback–I: No bandwidth constraint. *IEEE Transactions on Information Theory*, 12(2):172–182, 1966.
[28] C. E. Shannon. The zero error capacity of a noisy channel. *Information Theory, IEEE Transactions on*, 2(3):8–19, 1956.
[29] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp. Lower bounds to error probability for coding on discrete memoryless channels. *Information and Control*, 10(1):65–103, 1967.
[30] İ. E. Telatar. *Multi-Access Communications with Decision Feedback Decoding*. Ph.d. thesis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, May 1992.
[31] İ. E. Telatar and R. G. Gallager. New exponential upper bounds to error and erasure probabilities. In *ISIT 1994, Trondheim, Norway June 27- July 1, 1994*, 1994.
[32] H. Yamamoto and K. Itoh. Asymptotic performance of a modified Schalkwijk-Barron scheme for channels with noiseless feedback. *Information Theory, IEEE Transactions on*, 25(6):729–733, 1979.
[33] K. Sh. Zigangirov. Upper bounds for the error probability for channels with feedback. *Problems of Information Transmission*, 6(2):87–92, 1970.
[34] K. Sh. Zigangirov. Optimum zero rate date transmission through binary symmetric channel with feedback. *Problems of Control and Information Theory*, 7(3):21–35, 1978.